

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using 6w model

Run on: February 27, 2006, 06:33:05 ; Search time 6805.28 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182

Sequence: 1 atggcgctgcttaccattatc.....aggttaagcagctcgtaagc 1182

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmb1:*
1: gb_da:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_str:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1171.2	99.1	1182	6	ES0425
2	1171.2	99.1	2775	1	DQ086153
3	1171.2	99.1	3045	6	ES0424
4	1169.6	99.0	4315	1	AB039950
5	1166.4	98.7	2989	1	AB016245
6	1166.4	98.7	200690	1	BSUB0019
7	1164.8	98.5	3738	1	AB046355
8	1056.4	89.4	8189	1	BSZ92954
9	772.8	65.4	110000	1	AE017331_36
10	772.8	65.4	110000	1	CP000002_36
11	733.2	62.0	8164	1	BSZ92953
12	548.4	46.4	819	6	AX434279
13	507.2	42.9	3244	1	BACCAPAC
14	507.2	42.9	94829	1	AE011191
15	507.2	42.9	94830	1	AE017335
16	507.2	42.9	96231	1	AF188935
17	430	36.4	110000	1	BA000028_02
18	397	33.6	1336	1	BF060883

19	338.6	28.6	3240	1	AF269703	AF269703 Staphyloc
20	338.6	28.6	3240	6	AR485657	AR485657 Sequence
21	338.6	28.6	3240	6	AX145021	AX145021 Sequence
22	338.6	28.6	110000	1	CP000029_21	Continuation (22 o
23	338.6	28.6	300275	1	AE016751	AE016751 Staphyloc
24	324.4	27.4	110000	1	AP006716_05	Continuation (6 of
25	323.6	27.4	110000	1	AP008934_03	Continuation (4 of
26	246.2	20.8	861	6	AR484742	AR484742 Sequence
27	246.2	20.8	861	6	AX143599	AX143599 Sequence
28	181	15.3	309	6	AX434752	Continuation (9 of
29	160.2	13.6	110000	1	AJ749949_08	Continuation (16 o
30	134.4	11.4	110000	1	AE009951_15	Continuation (26 o
31	127.4	10.8	110000	1	AE017340_25	Continuation (26 o
32	113.6	9.6	304450	1	BX294134	BX294134 Pirellula
33	88.8	7.5	301508	1	AE017292	AE017292 Leptospir
34	85.6	7.2	10771	1	AE011422	AE011422 Leptospir
35	58.2	4.9	7218	6	166494	166494 Sequence 14
36	52.6	4.5	2000	6	AX655393	AX655393 Sequence
37	45.4	3.8	2352	6	AR547539	AR547539 Sequence
38	45.4	3.8	9109	2	CFR09D5	AL023812 Caenorhab
39	44.6	3.8	125020	8	AR429315	AF429315 Homo sapi
40	44.4	3.8	10437	1	AE006060	AE006060 Pasteurel
41	44.4	3.7	199901	14	AC154148	AC154148 Muntiacus
42	43.8	3.7	171218	14	AC156153	AC156153 Bos tauru
43	43.6	3.7	250029	2	AE014820	AE014820 Plasmodiu
44	43.2	3.7	134213	8	CNS05TEV	AL359399 Human chr
45	42.8	3.6	1818	1	AF267210	AF267210 Candidatu

ALIGNMENTS

RESULT 1	ES0425	1182 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	ES0425	Process for producing poly-gamma-glutamic acid.			
DEFINITION	ES0425	UP 2001017182-A/2.			
ACCESSION	ES0425	UP 2001017182-A/2.			
VERSION	ES0425	UP 2001017182-A/2.			
KEYWORDS	UP 2001017182-A/2.				
SOURCE	Bacillus subtilis				
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
REFERENCE	1 (bases 1 to 1182)				
AUTHORS	Ashuchi, M., Misono, H. and Soda, K.				
TITLE	Process for producing poly-gamma-glutamic acid				
JOURNAL	Patent: JP 2001017182-A 2 23-JAN-2001;				
COMMENT	NAGASE & CO LTD				
OS	Bacillus subtilis (hay bacillus) IFO 3336				
PN	JP 2001017182-A/2				
PD	23-JAN-2001				
PR	09-JUL-1999 JP 1999196335				
PI	MAKOTO ASHUCHI, HARUO MISONO, KENJI SODA				
PC	CI2N15/09, COB66/36, CI2N1/15, CI2N1/19, CI2N1/21, CI2N5/10, PC				
CC	CI2N9/00, CI2N9/90				
PH	CI2P13/14, CI2N15/00, CI2N5/00				
FT	Key	Location/Qualifiers			
FT	source	1..1182			
FT		/organism='Bacillus subtilis (hay bacillus)'			
FT		IFO 3336'			
FEATURES	source	Location/Qualifiers			
		1..1182			
		/organism='Bacillus subtilis'			
		/mol_type='genomic DNA'			
		/db_xref='taxon:1423'			
ORIGIN					
Query Match	99.1%;	Score 1171.2;	DB 6;	Length 1182;	
Best Local Similarity	99.7%;	Pred. No. 5.1e-250;			
Matches 1173;	Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;
7	TGTTACTCATTTATAGCGTGTGCTGTCATACGATCGATCGAATATTAGAAAAACGACGA	66			

|||||
4 TGGTACTCATATTATGCTGTGTGTCTATACCTGGATCATCGGAATTTAGAAAAAGACGA 63
67 CATCAAAAAACATTGATGCCCCCTCTGTTGGGTGAATATTACGGCATCCGCGAAAA 126
64 CATCAAAAAACATTGATGCCCCCTCTGTTGGGTGAATATTACGGCATCCGCGAAAA 123
127 TCGACTGTGACAAAGCTGACAAACCGGAATATTATAGAAAGCCGGTTACAGACTTTGGA 186
124 TCGACTGTGACAAAGCTGACAAACCGGAATATTATAGAAAGCCGGTTACAGACTTTGGA 183
187 AAAACAAGAGAAACAGATGACAAAGATTTACTGGGACACACCGGAGAAAAAGCCGATT 246
184 AAAACAAGAGAAACAGATGACAAAGATTTACTGGGACACACCGGAGAAAAAGCCGATT 243
247 AAAAGGAAACCTCAGGGCCCGAATATCGAGAGCAAAAAAGAAAGTCATGAGAGAAACGTA 306
244 AAAAGGAAACCTCAGGGCCCGAATATCGAGAGCAAAAAAGAAAGTCATGAGAGAAACGTA 303
307 AAAAGGAGGGCTAAAGCGATTTGTCAATGATGAGCGCTTAAACCGCATTTATCAATC 366
304 AAAAGGAGGGCTAAAGCGATTTGTCAATGATGAGCGCTTAAACCGCATTTATCAATC 363
367 ATCTTTCAAGAAAGAACTTCTGAGAGCCAAATATCGGCGTCATTTGTGAATGTTTAGAAGAC 426
364 ATCTTTCAAGAAAGAACTTCTGAGAGCCAAATATCGGCGTCATTTGTGAATGTTTAGAAGAC 423
427 CATATGATGATGATGAGGGCCGACGCTTGATGAAATTTGCAAGAGCTTTACCGCTCAAT 486
424 CATATGATGATGATGAGGGCCGACGCTTGATGAAATTTGCAAGAGCTTTACCGCTCAAT 483
487 CCTTAAATGAGCACTTGTCAATACATGATGATGATGATGATGATGATGATGATGATGATGAT 546
484 CCTTAAATGAGCACTTGTCAATACATGATGATGATGATGATGATGATGATGATGATGATGAT 543
547 GCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 606
544 GCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 603
607 TTACGTAATTTGAAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
604 TTACGTAATTTGAAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663
667 CAAGAGCTCGGATGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 726
664 CAAGAGCTCGGATGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 723
727 CCGGAGAGAAATGAGAAATTTCTTCGCTGATCAGTCCGAGGAGGCTGGGCACTTTGTTAAT 786
724 CCGGAGAGAAATGAGAAATTTCTTCGCTGATCAGTCCGAGGAGGCTGGGCACTTTGTTAAT 783
787 GGGTTTGGCGGAAAGAGAGCTTTCTTACTTTGAAATATATGAAAGCTGTAAAGAAATC 846
784 GGGTTTGGCGGAAAGAGAGCTTTCTTACTTTGAAATATATGAAAGCTGTAAAGAAATC 843
847 GGTATACCGGAGAGAGATCCGATCATCATGAACTGCGCGGAGAGAGAGAGAGAGAGAGAGAG 906
844 GGTATACCGGAGAGAGATCCGATCATCATGAACTGCGCGGAGAGAGAGAGAGAGAGAGAGAG 903
907 ACACAGCAATTTGCAAAATGACGTAATTTGCTTAAATTTGAAAGAAAGTGAATGATCTTAATC 966
904 ACACAGCAATTTGCAAAATGACGTAATTTGCTTAAATTTGAAAGAAAGTGAATGATCTTAATC 963
967 GGTGAAACAAACAGAACCGATCGTAAAGAGCTTATGAAAGAGCAAAATTCCTGACAGCAAA 1026
964 GGTGAAACAAACAGAACCGATCGTAAAGAGCTTATGAAAGAGCAAAATTCCTGACAGCAAA 1023
1027 CTGCAATGACTTGAATTAAGTCAACAGATGAAATTTATGAAATTTGTTAAAGAAAGAAATG 1086
1024 CTGCAATGACTTGAATTAAGTCAACAGATGAAATTTATGAAATTTGTTAAAGAAAGAAATG 1083
1087 CACAACCGGTGATATATGAGCGGCAATATTTCAATGAGCGGAGAGAGCTTTAATTTGAA 1146

Db 1084 CACAACCGGTGATATATGAGCGGCAATATTTCAATGAGCGGAGAGAGCTTTAATTTGAA 1143
Oy 1147 AAATCCAGCAATCAAGGTAAAGCAGCTGTAAGC 1182
Db 1144 AAATCCAGCAATCAAGGTAAAGCAGCTGTAAGC 1179
RESULT 2
LOCUS DQ086153 2775 bp DNA linear BCT 27-JUN-2005
DEFINITION Bacillus subtilis strain ZU-7 pgab (pgab), pgac (pgac), and pgaa
ACCESSION DQ086153
VERSION 1 (bases 1 to 2775)
KEYWORDS DQ086153.1 GI:68138277
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
REFERENCE Shi, F., Xu, Z., and Cen, P.
AUTHORS Shi, F., Xu, Z., and Cen, P.
TITLE Efficient production of poly(glutamic acid) by a new strain
JOURNAL Bacillus subtilis ZU-7
Unpublished
REFERENCE 2 (bases 1 to 2775)
AUTHORS Shi, F., Xu, Z., and Cen, P.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-2005) Institute of Bioengineering, Department of
Chemical Engineering and Bioengineering, Zhejiang University, No.38,
Zheda Road, Hangzhou, Zhejiang 310027, P.R. China
FEATURES
source
1..2775
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="ZU-7"
/db_xref="taxon:1423"
1..1182
/gene="pgab"
1..1182
/gene="pgab"
/function="biosynthesis of poly(glutamic acid)"
/codon_start=1
/transl_table=1
/product="Pgab"
/protein_id="AAV85940.1"
/db_xref="gi:68138278"
/translation="MMLIIICAVIIVYIGIEKRRKXNDALPRVNINGIRKSTV
TRLTTLGILENGYKTVGKTGTDAKMTYMDPEKPKRKQGNIGGQKVEETVE
RGANAIVSECAVNPDIQIIFQELQANIGVNVVEDHNDVNGPTLDELAEAFAT
IPVNGHIVITDSEYTERFKQAKERNKTVIADNSKITDEYLRFKEVAVFPDNLAL
GVAQALGIDETAFKGMUNAPPDGAMRIILPLISPSRPHVNGFAANDASSTINIK
RVKEIGYPTDPLIINNCRADRVDRTOGFANDVLPYIASESLIIGETTERIVKAYE
GKTPADKLHDELFTSTDEINELKRRKRNRYIVGVNIGHAEPLEIKIHRYKQQLV
S"
1183..1632
/gene="pgac"
1183..1632
/gene="pgac"
/function="biosynthesis of poly(glutamic acid)"
/codon_start=1
/transl_table=1
/product="Pgac"
/protein_id="AAV85941.1"
/db_xref="gi:68138279"
/translation="MFGSDLIYALILGVLSLIRAEKTIYPAQLVVPYIGLVNQP
AVFLLVGLIVPLLTIVIVYKSKPMIYGRKRFAMILITGLIADDFLPVIFEEI
AEFRIGIIVPLIANTIOKQGLITITGSLILSGAIFALMFVYLLI"
1633..2775
/gene="pgaa"
1633..2775
/gene="pgaa"
/function="biosynthesis of poly(glutamic acid)"
/codon_start=1
/transl_table=1


```

/product="PgaA"
/protein_id="AAV85942.1"
/db_xref="GI:68138280"
/translation="MKKLSPEKLLKLTQOKKTKNKVFIAPDIVVLMFAFWAG
KAETPKVTSVDVLSASFEVDIMMRVKEVELEOKGDSIFQYVEPIFRASDVAGN
FENPVTKYKQADKEIHLQTNKESVKLKMFTVNSANHEMDYGVOKMDTLG
EPKONLDIVAGVSLDAKKISYOKNGVITATLGFSDYKGFPAKKKPTGVLPA
DPEIPIVISEKKKADIVVVSHGQSDNDPNRQOQLARMSDAGADIIIVGHHP
VLEPIEVNGVIFYSLSGFVFDQMTIRDSALVQYHKKNGRFEVETPIIDHEAT
PAPVKDLSIKQITIIRELTKDSNFAMKVEDKLTLPDIDHSLKSK"

```

ORIGIN

```

Query Match 99.1%; Score 1171.2; DB 1; Length 2775;
Best Local Similarity 99.7%; Pred. No. 4.4e-250;
Matches 1173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

127 TCGACTGTGACAAAGCTGCAACCCGGAATATTAAGCCGGTTACAAAGCTGTGGA 186
124 TCGACTGTGACAAAGCTGCAACCCGGAATATTAAGCCGGTTACAAAGCTGTGGA 183
187 AAAACAACAGAAACAGATGCAAGATGATTTACTGGAGACACCCGAGGAAAAAGCCGATT 246
184 AAAACAACAGAAACAGATGCAAGATGATTTACTGGAGACACCCGAGGAAAAAGCCGATT 243
247 AAAACGAAACCTCAGGGGCGGAATATCGAAGCAAAAAGAACTCATGAGAGAAACAGTA 306
244 AAAACGAAACCTCAGGGGCGGAATATCGAAGCAAAAAGAACTCATGAGAGAAACAGTA 303
307 GAAAGAGGGGCTTAACCGGATTTGTCAGTGAATGATGGCTTTTAAACCAATTC 366
304 GAAAGAGGGGCTTAACCGGATTTGTCAGTGAATGATGGCTTTTAAACCAATTC 363
367 ATCTTTGAGGAAAGATCTTGGAGGCGCAATATGGGGCTATGTAATGTTTAAAGAAC 426
364 ATCTTTGAGGAAAGATCTTGGAGGCGCAATATGGGGCTATGTAATGTTTAAAGAAC 423
427 CATATGATGTCAATGGGCGCGAGCTTGAATGAATTCAGAAAGCGTTTACCGTACAAAT 486
424 CATATGATGTCAATGGGCGCGAGCTTGAATGAATTCAGAAAGCGTTTACCGTACAAAT 483
487 CCTTATTAATGGCAATCTTGTGATTAAGATGATATACCGAGTTCTTTAAACAAAA 546
484 CCTTATTAATGGCAATCTTGTGATTAAGATGATATACCGAGTTCTTTAAACAAAA 543
547 GCAAAAAGAGCAAAACAAAGATCATATGCTGATTAACCAAAAATTACAGATGAGTAT 606
544 GCAAAAAGAGCAAAACAAAGATCATATGCTGATTAACCAAAAATTACAGATGAGTAT 603
607 TTAAGTAAATTTGAATACATGATATTCCTGATTAACCGTTCTCGGCGCTGGGTGGCT 666
604 TTAAGTAAATTTGAATACATGATATTCCTGATTAACCGTTCTCGGCGCTGGGTGGCT 663
667 CAAGCACTCGGCAATGACGAAGAAACAGCAATTTAAGGAATGCTGATGCGCGCGAGAT 726
664 CAAGCACTCGGCAATGACGAAGAAACAGCAATTTAAGGAATGCTGATGCGCGCGAGAT 723
727 CCGGAGCAATGAGAAATCTTCGCTGATGAGTCCGAGCGAGCGCTGGGCACTTTGTAAT 786
724 CCGGAGCAATGAGAAATCTTCGCTGATGAGTCCGAGCGAGCGCTGGGCACTTTGTAAT 783
787 GGGTTTCCGCAACAGAGCTTCTTCTACTTTGAATATGAAACCGTGTAAAGAAATC 846
784 GGGTTTCCGCAACAGAGCTTCTTCTACTTTGAATATGAAACCGTGTAAAGAAATC 843
847 GGTTAACCGCAAGATGATCCGATCATATGAATTCGCCGCGAGACCGTGTGATCGG 906

```

```

Db 844 GGTTAACCGCAAGATGATCCGATCATATGAATTCGCCGCGAGACCGTGTGATCGG 903
Qy ACACAGCAATTCGCAAAATGAGTATTCGCTTATTTGAAGCAAGTGAATGATCTTAATC 966
Db ACACAGCAATTCGCAAAATGAGTATTCGCTTATTTGAAGCAAGTGAATGATCTTAATC 963
Qy GGTGAACCAAGCAACCGATGCTAAAGCCATGAGAGCAAAATTCCTGCAGACAAA 1026
Db GGTGAACCAAGCAACCGATGCTAAAGCCATGAGAGCAAAATTCCTGCAGACAAA 1023
Qy CTGCATGACCTAGATATTAAGTCAACAGATGAATTTATGAAATTTTAAAGAAATG 1086
Db CTGCATGATCTAGATATTAAGTCAACAGATGAATTTATGAAATTTTAAAGAAATG 1083
Qy CACAACCGTGTCAATATGAGGCTCGGCAATATTCATGTTGCGGAGACCTTTAATTGA 1146
Db CACAACCGTGTCAATATGAGGCTCGGCAATATTCATGTTGCGGAGACCTTTAATTGA 1143
Qy AAAATCCAGAAATACAGATTAAGAGAGCTCGTAAGC 1182
Db AAAATCCAGAAATACAGATTAAGAGAGCTCGTAAGC 1179

```

```

RESULT 3
E50424 3045 bp DNA linear PAT 31-JAN-2002
LOCUS E50424
DEFINITION Process for producing poly-gamma-glutamic acid.
ACCESSION E50424
VERSION E50424.1 GI:18629412
KEYWORDS JP 2001017182-A/1.
SOURCE JP 2001017182-A/1.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 3045)
AUTHORS Ashiuchi, M., Misono, H. and Soda, K.
TITLE Process for producing poly-gamma-glutamic acid
JOURNAL Patent: JP 2001017182-A 1 23-JAN-2001;
NAGASE & CO LTD
COMMENT OS Bacillus subtilis (hay bacillus) IFO 3336
PN JP 2001017182-A/1
PD 23-JAN-2001
PF 09-JUL-1999 JP 1999196335
PR
PT MAKOTO ASHIUCHI, HARUO MISONO, KENJI SODA
PC C12N15/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
C12N9/00, C12N9/90.
PC C12P13/14, C12N15/00, C12N5/00
CC
FH Key 1.3047 Location/Qualifiers
FT source 1.3045 /organism="Bacillus subtilis"
FT IFO 3336. /mol_type="genomic DNA"
FT /db_xref="taxon:1423"

```

```

FEATURES
source 1.3045 Location/Qualifiers
1.3045 /organism="Bacillus subtilis"
/mol_type="genomic DNA"
/db_xref="taxon:1423"

```

ORIGIN

```

Query Match 99.1%; Score 1171.2; DB 6; Length 3045;
Best Local Similarity 99.7%; Pred. No. 4.4e-250;
Matches 1173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

7 TGGTACTCATTAATACCTGTGCTGTGCATATGCTCATGCAATATTTAGAAAAAGACGA 66
Db TGGTACTCATTAATACCTGTGCTGTGCATATGCTCATGCAATATTTAGAAAAAGACGA 81
Qy CATCAGAAAAACATTGATGCTCCCTCTGTTGGGTGAATTTAAACGCAATCCGCGGAAAA 126
Db CATCAGAAAAACATTGATGCTCCCTCTGTTGGGTGAATTTAAACGCAATCCGCGGAAAA 141
127 TCGACTGTGACAAAGCTGCAACCCGGAATATTAAGCCGGTTACAAAGCTGTGGA 186

```

Db 142 TCGATGTGACAAAGCTGACAAACCGGAATATTAAAGAACCGGTACTCAAGCTGTGGA 201
Oy 187 AAAACAACAGGAACGATGCAAGAAATGATTTACTGGACACACCGGAGGAAAACCGCAT 246
Db 202 AAAACAACAGGAACGATGCAAGAAATGATTTACTGGACACACCGGAGGAAAACCGCAT 261
Oy 247 AAAAGGAACCTCAGGGGCGCAATTCGGAGAGCAAAAAGAAAGTATGAGAAACAGTA 306
Db 262 AAAAGGAACCTCAGGGGCGCAATTCGGAGAGCAAAAAGAAAGTATGAGAAACAGTA 321
Oy 307 GAAAGAGGGGCTTACGCGATTTGTCAGTGAATGTCATGGCTGTATACCGAGATTATCAATC 366
Db 322 GAAAGAGGGGCTTACGCGATTTGTCAGTGAATGTCATGGCTGTATACCGAGATTATCAATC 381
Oy 367 ATCTTTACAGGAAGAACTTTGTCAGGCGCAATATCGCGCTCATTTGTGAATGTTTAAAGAC 426
Db 382 ATCTTTACAGGAAGAACTTTGTCAGGCGCAATATCGCGCTCATTTGTGAATGTTTAAAGAC 441
Oy 427 CATATGAGATGTATGGGGCGGCGCTTGATGAAATTTGCAAGAGCGTTTACCGCTCAAT 486
Db 442 CATATGAGATGTATGGGGCGGCGCTTGATGAAATTTGCAAGAGCGTTTACCGCTCAAT 501
Oy 487 CCTTATTAATGGCCATCTTGTCTATTAAGATAGTGAATATACCGAGTTCTTTAAACAAA 546
Db 502 CCTTATTAATGGCCATCTTGTCTATTAAGATAGTGAATATACCGAGTTCTTTAAACAAA 561
Oy 547 GCAAAAAGAACGAAACACAAAGATCATCTTGTCTGATCTCAAAAATTACAGATGAGT 606
Db 562 GCAAAAAGAACGAAACACAAAGATCATCTTGTCTGATCTCAAAAATTACAGATGAGT 621
Oy 607 TTACGTAATTTGATATGATGATGATTCCTGATTAACGCTTCTGCGCTGGGTGGCT 666
Db 622 TTACGTAATTTGATATGATGATGATTCCTGATTAACGCTTCTGCGCTGGGTGGCT 681
Oy 667 CAAGCACTCGGCACTTGTGACGAAGAAACAGCATTTAAAGGAATCTGATGCGCGCAGAT 726
Db 682 CAAGCACTCGGCACTTGTGACGAAGAAACAGCATTTAAAGGAATCTGATGCGCGCAGAT 741
Oy 727 CCGGAGCAATGAAATTTCTTCGCTGATCACTCGGAGCGCTGGGCACTTTGTTAAT 786
Db 742 CCGGAGCAATGAAATTTCTTCGCTGATCACTCGGAGCGCTGGGCACTTTGTTAAT 801
Oy 787 GGGTTGCGCGCAAGGAGCGCTTCTTACTTTGAATATATGAAGCGTGAAGAAATC 846
Db 802 GGGTTGCGCGCAAGGAGCGCTTCTTACTTTGAATATATGAAGCGTGAAGAAATC 861
Oy 847 GGTACCCGAGCGATGATCGATCATCATGAACTGCGCGCAGACCGGTGCGATCGG 906
Db 862 GGTACCCGAGCGATGATCGATCATCATGAACTGCGCGCAGACCGGTGCGATCGG 921
Oy 907 ACAACAGCAATTTGGCAATGACGTAATTCCTTATTTAAAGCAAGTGAATCTGATCTTAATC 966
Db 922 ACAACAGCAATTTGGCAATGACGTAATTCCTTATTTAAAGCAAGTGAATCTGATCTTAATC 981
Oy 967 GGTGAAGCAAGAACCGATGTAAGAGCTATGAAGAGGCAAAATTCCTGACAGCAA 1026
Db 982 GGTGAAGCAAGAACCGATGTAAGAGCTATGAAGAGGCAAAATTCCTGACAGCAA 1041
Oy 1027 CTGCATGACCTAGATGATTAAGTCAACAGATGAAATTTGTTAAAGAAAGATG 1086
Db 1042 CTGCATGACCTAGATGATTAAGTCAACAGATGAAATTTGTTAAAGAAAGATG 1101
Oy 1087 CACAACCGGTGATATATATGCGCTGCGCAATTTATCATGTGCGCGACAGCTTTAATTGAA 1146
Db 1102 CACAACCGGTGATATATATGCGCTGCGCAATTTATCATGTGCGCGACAGCTTTAATTGAA 1161
Oy 1147 AAAATTCACGAATACAAAGTAAAGCGCTCGTAAGC 1182
Db 1162 AAAATTCACGAATACAAAGTAAAGCGCTCGTAAGC 1197

, RESULT 4

AB039950
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
1. 4315
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="Aeahi:kawa"
/db_xref="taxon:1423"
941..2122
/gene="capB"
941..2122
/gene="capB"
/codon_start=1
/transl_table=1
/product="capB"
/protein_id="BAB13484.1"
/db_xref="GI:10119861"
/translation="MMLLIACAVILVIGILEKRRHOKNTDALPVRVNINGIRKSTV
TRLTGILIEAGYKVGKTGTSDAMTYMDPEEKPIKPKQGNIGQKEMETVE
RGANAIVSECAVNDVVOIIFOEELNANGIVNVFEDHMDVNGPTLDEIAEAFAT
IPNGHLYITSEYTEFPKQAKERNTKVIIANDSKITDEYLRKEFVNPFPDMSIAL
GVAQALGIDESTARKMLNAPDPGAKHILPLISPSEBGEVYVNGPANDASSTINIK
RVKIKGYPTDPIIIMNCRADRVDTQFADVLPVLEASLLIGETETPVAYEE
GKIPADKLHDLREKSTDEIMELKKRMHNRVIYGVNIGHAEBPLIEKHIEYKVKOLV
S"
2137..2586
/gene="capC"
2137..2586
/gene="capC"
/codon_start=1
/transl_table=1
/product="capC"
/protein_id="BAB13485.1"
/db_xref="GI:10119862"
/translation="MFGSDLYIALILGVLSLIFAEKGTIVPAGLVVPVIGLVNOR
VFILVILVSLITVIVIKYGLSKFMILIGRRKFPAMITGVLAIFDPLVIVPEI
AEFRIGIIVGLIANTIOKQGLITTFGSTILLSGATPAIMFVYLL"
2605..3747
/gene="capA"
2605..3747
/gene="capA"
/codon_start=1
/transl_table=1
/product="capA"
/protein_id="BAB13486.1"
/db_xref="GI:10119863"
/translation="MKKELSFHEKLLKLTQOKKTKNGVEFALPIVFLMFAEWAG
KAEPKVKYSDVLSAFVGDIMNGRYVEKVTGKGADSIFOYVEPIFRASDVAGN
FENPVYOKNROADKEITHLOTNKESVVLKDNMFTVNSANNMADYGVGMDVTLG
EPKRONDIVAGSYLSDAKKKISYOKKNGTITATIGETDYSGGGFAKKNTPGVLP
DPEITPWIISAKKHADIIVVQSHMGDEYDNDPDRQQLRAHNSDAGADIIIVGHNP
VLEPIEVNGTVIFISLGNFVFDQGTWTRDSALVQYHLKNGGTGRFEVTPIDHEAT
PAPVKDSLKOKTIIIRELTKDSNPFAMKVEDKLTFFDIDHSKLSK"
3762..3929
gene

CDS
1197..1646
/gene="pgsc"
/codon_start=1
/transl_table=11
/product="pgsc"
/protein_id="BAA85264.1"
/db_xref="GI:6045073"
/translation="MFSGDLYIALILGVLSLIFAEKTIIVPAGLVVPGVGLVFNOP
VFILVLVLLTYIVYKVGSKFMILYGRKFAALITGLVLAFAFPLVPIVPEI
AEFGIGIIVPGLANTIQKQGLTTPGSTILLSGATPAIMFVYLLI"
1665..2807
/gene="pgsa"
1665..2807
/gene="pgsa"
/codon_start=1
/transl_table=11
/product="pgsa"
/protein_id="BAA85265.1"
/db_xref="GI:6045074"
/translation="MKELSPHEKLLKTKOOKKKTNHVEIALPIVEVLFAMFAMAG
KAEPKVTYSDVLSASFVDIMMGRVVEYEBOKGADSI FOYEPIFRASDYVGN
FENPTVOKYKQADKEIHLCTNKESVULKDMPTVLSANNHAMDPYOGMDTIG
EPAKONLDIVAGHSLSDAKKISTYQKNGVTIATLFTDVSGKFAKKNPGLPLA
DPELFIEMIISAQRADIVVQSHWGQBYDNDPDRQLARASMDAGADIIVGHHPH
VLEPIEVNGTVIFYSIGNFVFDQWTRDSALVOYHLKKNGTGRFEVTPIDTHEAT
PAPVKDLSIKOTIIRLETKDSNFAMKVEDKLTFFIDHSKLSKX"

ORIGIN

Query Match 98.7%; Score 1166.4; DB 1; Length 2989;
Best Local Similarity 99.5%; Pred. No. 5.1e-249;
Matches 1170; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 TGGTACTCATATATGACCTGTGCTGTGCTGATCATCGGAATATTAGAAAAAGACGA 66
DB 4 TGGTACTCATATATGACCTGTGCTGTGCTGATCATCGGAATATTAGAAAAAGACGA 63
QY 67 CATCAGAAAAACATTGATGCCCTCCCTGTCGGGTGAATATTAAACGCAATCCGCGAAAA 126
DB 64 CATCAGAAAAACATTGATGCCCTCCCTGTCGGGTGAATATTAAACGCAATCCGCGAAAA 123
QY 127 TCGACTGTGACAAAGCTGACAAACCGGAATATTAAATGAAAGCCGTTACAACTGTTGCA 186
DB 124 TCGACTGTGACAAAGCTGACAAACCGGAATATTAAATGAAAGCCGTTACAACTGTTGCA 183
QY 187 AAAACAAGAGAAAGATGCAAGATGATTTAATCTGGGACACACCGGAGGAAAAAGCCGATT 246
DB 184 AAAACAAGAGAAAGATGCAAGATGATTTAATCTGGGACACACCGGAGGAAAAAGCCGATT 243
QY 247 AACCGGAAACCTCAGGGCCGGAATATCGAGAGCAAAAAAGAGTANTGAGAGAAACAGTA 306
DB 244 AACCGGAAACCTCAGGGCCGGAATATCGAGAGCAAAAAAGAGTANTGAGAGAAACAGTA 303
QY 307 GAAAGAGGGGCTTAACCGCATTTCTCAGTGAATCATAGGCTGTTAAACCCAGATTATCAATC 366
DB 304 GAAAGAGGGGCTTAACCGCATTTCTCAGTGAATCATAGGCTGTTAAACCCAGATTATCAATC 363
QY 367 ATCTTTCAGAGAACTCTCAGGCGCAATATCGCGCATTTGTAAGTTTAAAGAC 426
DB 364 ATCTTTCAGAGAACTCTCAGGCGCAATATCGCGCATTTGTAAGTTTAAAGAC 423
QY 427 CATATGATGTCATGGGGCCGAGCTTGATGAATTTGCAAGAGCGTTTACCGCTCAAT 486
DB 424 CATATGATGTCATGGGGCCGAGCTTGATGAATTTGCAAGAGCGTTTACCGCTCAAT 483
QY 487 CCTTATATGCGCATCTTGTCTATTACAGATAGTGAATATCCGAGTTCTTTTAAACAAAA 546
DB 484 CCTTATATGCGCATCTTGTCTATTACAGATAGTGAATATCCGAGTTCTTTTAAACAAAA 543
QY 547 GCAAAAGAACGAAACAAAGATCATCTGCTGATTAACCTCAAAATTTACAGATAGAT 606
DB 544 GCAAAAGAACGAAACAAAGATCATCTGCTGATTAACCTCAAAATTTACAGATAGAT 603
QY 607 TTACGTAATTTTGAATATCATGATTTCCCTGATTAACGCTTCTCTGCGCGCTGGTGGCT 666

DB 604 TTACGTAATTTTGAATATCATGATTTCCCTGATTAACGCTTCTCTGCGCGCTGGTGGCT 663
QY 667 CAAGACCTCGGCATTTGCAAGAAAGACATTTTAAAGGAATGCTGAATGCGCCGACAGAT 726
DB 664 CAAGACCTCGGCATTTGCAAGAAAGACATTTTAAAGGAATGCTGAATGCGCCGACAGAT 723
QY 727 CCGGAGCAATGAGAAATTTCTTCCGCTGATCATGCTCCGAGCGAGCTGGGCACTTTGTTAT 786
DB 724 CCGGAGCAATGAGAAATTTCTTCCGCTGATCATGCTCCGAGCGAGCTGGGCACTTTGTTAT 783
QY 787 GGGTTTGGCCGAAACGACGCTTCTTCTTCACTTTGAATATGAGAAACGTTTAAAGAAATC 846
DB 784 GGGTTTGGCCGAAACGACGCTTCTTCTTCACTTTGAATATGAGAAACGTTTAAAGAAATC 843
QY 847 GGTAAACCGACGAGTATCCGATCATCATGAACTCGCGCGAGACCGGTGCGATCGG 906
DB 844 GGTAAACCGACGAGTATCCGATCATCATGAACTCGCGCGAGACCGGTGCGATCGG 903
QY 907 ACACAGCAATTCGCAAAATGACGATATTCGCTTATATTGAAGCAATGAACTGATCTTATC 966
DB 904 ACACAGCAATTCGCAAAATGACGATATTCGCTTATATTGAAGCAATGAACTGATCTTATC 963
QY 967 GGTGAACACAGAACCGATGTTAAAGCCTTATGAAGAGCAAAATTTCTGCGACAGAA 1026
DB 964 GGTGAACACAGAACCGATGTTAAAGCCTTATGAAGAGCAAAATTTCTGCGACAGAA 1023
QY 1027 CTGCATGACCTAGAGTATTAAGTCAACAGATGAATTTAGAAATTTTAAAGAAAGATG 1086
DB 1024 CTGCATGACCTAGAGTATTAAGTCAACAGATGAATTTTAAAGAAAGATG 1083
QY 1087 CACAACCGTGTCAATATATGCGGTGCGCAATATTCATGCTGCGGAGACCTTTTATTTGAA 1146
DB 1084 CACAACCGTGTCAATATATGCGGTGCGCAATATTCATGCTGCGGAGACCTTTTATTTGAA 1143
QY 1147 AAAATCCAGAAATCAAGTTAAAGCAGCTGTAAGC 1182
DB 1144 AAAATCCAGAAATCAAGTTAAAGCAGCTGTAAGC 1179

RESULT 6
BSUB0019/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

Bacillus subtilis subsp. subtilis str. 168
Bacillus subtilis subsp. subtilis str. 168
to 3809670.
299122 AL009126
299122.2 GI:32468830

Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 200690)

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Beselieres, P., Bolognin, A., Borchert, S.,
Borries, R., Boursier, L., Brans, A., Braun, M., Briganti, S.C.,
Bron, S., Brouillette, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerth, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabis, C., Ferrari, E.,
Foliger, D., Fritz, S.C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Gilm, S.Y., Glaser, P., Goffeau, A., Golding, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Hennau, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portecelle, D., Portocarrero, A.M., Prescott, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,

Query Match	Best Local Similarity	98.7%;	Score 1166.4;	DB 1;	Length 200690;
Matches 1170;	Conservativity	99.5%;	Pred. No. 2.6e-249;	Mismatches 6;	Indels 0; Gaps 0;
Db	90671	TGGTACTCATTTATAGCCTGTGCTGTCTATCTGTGCATCGGAATTTAAGAAAAACGACGA	66		
Qy	67	CATCAGAAAAACATTGATGCGCCCTCCTGTTGGGTGAATATTAACGCGATCCGCGAAAA	126		
Db	90611	CATCAGAAAAACATTGATGCGCCCTCCTGTTGGGTGAATATTAACGCGATCCGCGAAAA	905522		
Qy	127	TGCACGTGTGCAAGGCGTGAACAACCGGAATTTAATAGAAGCCGCTTAACAAGACTTTGGA	186		
Db	90551	TCGACTGTGTGCAAGGCGTGAACAACCGGAATTTAATAGAAGCCGCTTTAACAAGACTTTGGA	904922		
Qy	187	AAAAACAACAGAAACAGATGCAAGAAATGTTAATCTGGGACACACCGGAGGAAAAAGCCGATT	246		
Db	90491	AAAAACAACAGAAACAGATGCAAGAAATGTTAATCTGGGACACACCGGAGGAAAAAGCCGATT	904322		
Qy	247	AAACGGAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAAAGTCATGAGAGAAACAGTA	306		
Db	90431	AAACGGAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAAAGTCATGAGAGAAACAGTA	903722		
Qy	307	GAAGAAGGGGCTAAACGCGATTGTCAGTGAATGCATGCGTGTAAACCCAGATTATCAATC	366		
Db	90371	GAAGAAGGGGCTAAACGCGATTGTCAGTGAATGCATGCGTGTAAACCCAGATTATCAATC	903122		
Qy	367	ATCTTTACAGAAAGAACTTCTGCAGGCCAATTCGGCGCTCATTTGGAAGTTTGAAGAC	426		
Db	90311	ATCTTTACAGAAAGAACTTCTGCAGGCCAATTCGGCGCTCATTTGGAAGTTTGAAGAC	902522		
Qy	427	CATATGATGTCAATGGGCGGACCGCTTGATGAATTTGAGAGAGCGTTTACCGCTCAATT	486		
Db	90251	CATATGATGTCAATGGGCGGACCGCTTGATGAATTTGAGAGAGCGTTTACCGCTCAATT	901922		
Qy	487	CCTTATTAATGAGCCATCTTGTCATTACAGATAGTGAAATTAACCGAGTCTTTAAACAAAA	546		
Db	90191	CCTTATTAATGAGCCATCTTGTCATTACAGATAGTGAAATTAACCGAGTCTTTAAACAAAA	901322		
Qy	547	GCAAAAAGAACGAAACACAAAAGTCATCTTCTGATTAACCAAAATTAACAGATAGAT	606		
Db	90131	GCAAAAAGAACGAAACACAAAAGTCATCTTCTGATTAACCAAAATTAACAGATAGAT	900722		
Qy	607	TTACGTATTTTGAATACATGGTATTCCTCTATTAACGCTTCTGCGCGCTGGGTGGCT	666		
Db	90071	TTACGTATTTTGAATACATGGTATTCCTCTATTAACGCTTCTGCGCGCTGGGTGGCT	900122		
Qy	667	CAAGCACTCGGCAATTGACGAAGAAACAGCATTTAAGGAATGCTGAATGCGCGCCAGAT	726		

Db	90011	CAAGCACTCGGCATGTAGCGAAGAAACAGATTTAAGGAAATGCTGAATGCGCCCGCAGAT	89955
Qy	727	CCGGGAGCAATGAGAAATTCCTCCGCTGATCAGTCCAGGAGCCTTGCGCACTTTGTTAAT	786
Db	89951	CCGGGAGCAATGAGAAATTCCTCCGCTGATCAGTCCGAGCAGCCTGGCACTTTGTTAAT	89899
Qy	787	GGGTTTCCCGCAACGACGGTTTCTTCTACTTAAATAATAGAAACGTGTAAAGAAATC	846
Db	89891	GGGTTTCCCGCAACGACGGTTTCTTCTACTTAAATAATAGAAACGTGTAAAGAAATC	89893
Qy	847	GGTACCCGACCGATGATCCGATCATCATATGAACTGCCGCGACGCTGTGCATCGG	906
Db	89931	GGTTACCCGACCGATGATCCGATCATCATATGAACTGCCGCGACGCTGTGCATCGG	89777
Qy	907	ACACAGCAATTCGCAATGACGTAATTCCTTAATTAAGCAAGTGAATGATCTTAATC	966
Db	89771	ACACAGCAATTCGCAATGACGTAATTCCTTAATTAAGCAAGTGAATGATCTTAATC	89711
Qy	967	GGTAAACCAACAGAACCGATCTGTAAAGCCTTAAGAAAGGCAAAATTCCTGCAGACAA	1026
Db	89711	GGTAAACCAACAGAACCGATCTGTAAAGCCTTAAGAAAGGCAAAATTCCTGCAGACAA	89655
Qy	1027	CTGCATGACCTAGAGTAAAGTCAACAGATGAATTAATGAAATTGTTAAAGAAAGATG	1086
Db	89651	CTGCATGATCTAGAGTAAAGTCAACAGATGAATTAATGAAATTGTTAAAGAAAGATG	89599
Qy	1087	CACAAACGGTGTCAATATATGCGCTGCGCAATATTCATGTCGCGCAGACCTTAATGAA	1146
Db	89591	CACAAACGGTGTCAATATATGCGCTGCGCAATATTCATGTCGCGCAGACCTTAATGAA	89533
Qy	1147	AAAAATCCACGATTCAGAGTTAAAGCAGCTGTAAAGC	1182
Db	89531	AAAAATCCACGATTCAGAGTTAAAGCAGCTGTAAAGC	89496
RESULT 7			
AB046355			
LOCUS	AB046355	3738 bp	DNA linear BCT 11-JAN-2002
DEFINITION	Bacillus subtilis ywec, ywCA, ywCB, ywTC genes, complete cds.		
ACCESSION	AB046355		
VERSION	AB046355.1	GI:13591556	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	1		
TITLE	Urushibata, Y., Tokuyama, S. and Tahara, Y.		
JOURNAL	Characterization of the Bacillus subtilis ywec gene, involved in		
PUBMED	gamma-polyglutamic acid production		
REFERENCE	J. Bacteriol. 184 (2), 337-343 (2002)		
AUTHORS	2 (bases 1 to 3738)		
TITLE	Tahara, Y. and Urushibata, Y.		
JOURNAL	Direct Submission		
PUBMED	Submitted (18-JUL-2000) Yasutaka Tahara, Shizuoka University,		
REFERENCE	Faculty of Agriculture, 836 Ohya, Shizuoka city, Shizuoka 422-8529,		
AUTHORS	Japan (E-mail: acyutah@aqr.shizuoka.ac.jp,		
TITLE	Tel: 81-54-238-4878 (ex. 7808), Fax: 81-54-237-3028)		
FEATURES			
source	Location/Qualifiers		
	1..3738		
	/organism="Bacillus subtilis"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:1423"		
	497..1678		
gene	/gene="ywec"		
	497..1678		
CDs	/gene="ywec"		
	/codon_start=1		
	/transl_table=1		
	/protein_id="BAB00948.1"		
	/db_xref="GI:13591557"		
	/translation="MWLLIACVAVLIVIGLEKRRHOKNIDALPRVNVINGIRKSTV		

TRUUTGILTEAGYKTKTCTGTADAMTYMDPEKKPIKRRKQGNIGEOKEVMEATVE
 RGAATVISECAANVPDYOIIRQEBLQANIVYVYEDHDWNGPFLDEIAEAFAT
 IPNGALVITDSEYTERFKQAKERNKVIITADSKITDEYLRFBETWPFNDASLAL
 GVAQALGIDEBEAFKMLNAPDPGNARKIILPLISPSBGFVNFPANDASTNATIK
 RKKEIGYPTDPIIIMNCRADRVDTQOFADVLPLYEASLILIGEOEPIVAYEE
 GKIPTADKLHDEYKSTDEIMELKKRHNRIYGVGNHGAEBPLIEKHIEYKVKQIV
 S"
 1693..2142
 /gene="ywcA"
 1693..2142
 /gene="ywcA"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB40949.1"
 /db_xref="GI:13591558"
 /translation="MFSGDYLIALILGVLISLIFAEKTIIVPAGLVVPGYLGVENOP
 VELLIVLIVLIVIVYKYGSKEMILYGRKFAMLLTGIVLKIAPFLPIVPEFI
 AEFPGITIVPGLIANTIQKQGLITTFSTILLSGRAPAIMFVYLI"
 2161..3303
 /gene="ywcB"
 2161..3303
 /gene="ywcB"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB40950.1"
 /db_xref="GI:13591559"
 /translation="MKKELSFHEKLLKTKQKKKTKNKHVIALPIVFLMFAEMWAG
 KAETPKVATSDVLSASFVGDIMGRVVEKTEQKADSIFQYVEPFRASDVYAGN
 PENPVTQKNDKEIHLQTNKESVYLKDMNFTVNSANNHMDYGVGMKDTLG
 EAPATNLDIVAGVSLDPAKKRISYQKNTGLATLGTVDYSGKFAKNTPGYLP
 DPEITPMISEAKKADIVVOSHMGQYDNDPRNQQLARMSDADADIIIVGHHP
 VLEPLEVNGTVIFSLGNFVFDQMTFRDSALVOYHLKKNQTRFEVTPIDTHEAT
 PAPVKDLSIKQTIIRLETKDSNFAMKVEDKLTPIIDHSDLSK"
 3318..3485
 /gene="ywcC"
 3318..3485
 /gene="ywcC"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB40951.1"
 /db_xref="GI:13591560"
 /translation="MKFVKAIWPFVAVALVFHMFSAFKFNQDLTDEKQKIDIMENKIKI
 QOQEBPVANK"
 ORIGIN
 Query Match 98.5%; Score 1164.8; DB 1; Length 3738;
 Beet Local Similarity 99.4%; Pred. No. 1.1e-248;
 Matches 1169; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

7 TGGTACTCATTAATAGCCTGTGCTGTCAATCTGGTCATCGAATATTAGAAAAACGACGA 66
 500 TGGTACTCATTAATAGCCTGTGCTGTCAATCTGGTCATCGAATATTAGAAAAACGACGA 559
 67 CATCGAAAAACATTGATGCTCTCTCTGTTGGGTGAATTTAAACGGCATCCCGCGAAAA 126
 560 CATCGAAAAACATTGATGCTCTCTCTGTTGGGTGAATTTAAACGGCATCCCGCGAAAA 619
 127 TCGACTGTGACAGGCTGACACCGGATATTATAGAAACCGGTACCAAGACTGTTGA 186
 620 TCGACTGTGACAGGCTGACACCGGATATTATAGAAACCGGTACCAAGACTGTTGA 679
 187 AAAACAAACGAAACAGATCAAGATGATTACTGGGACACACCGGAGGAAAAAGCGGATT 246
 680 AAAACAAACGAAACAGATCAAGATGATTACTGGGACACACCGGAGGAAAAAGCGGATT 739
 247 AAAACGAAACCTGAGGGGCGGATATTCGAGACCAAAAAAGACTCATGAGAAAGAAAGTA 306
 740 AAAACGAAACCTGAGGGGCGGATATTCGAGACCAAAAAAGACTCATGAGAAAGAAAGTA 799
 307 GAAAGAGGGGCTAACCGGATTGTCAGTGAATGCGATGGCTGTTAACCCAGATTATCAATC 366
 800 GAAAGAGGGGCTAACCGGATTGTCAGTGAATGCGATGGCTGTTAACCCAGATTATCAATC 859

367 ATCTTCAGAGAAAGAACTTCTGACAGGCCAATATCGGCTCATTTGTAATGTTTGAAGAC 426
 860 ATCTTCAGAGAAAGAACTTCTGACAGGCCAATATCGGCTCATTTGTAATGTTTGAAGAC 919
 427 CATATGATGTCATGAGGCGGACGCTTGATGAATTCAGAAAGGCTTTACCGCTACAAAT 486
 920 CATATGATGTCATGAGGCGGACGCTTGATGAATTCAGAAAGGCTTTACCGCTACAAAT 979
 487 CCTTATAATGGCCATCTTGTCATTAACAGATAGTAATACCGAGTCTTTTAAACAAAA 546
 980 CCTTATAATGGCCATCTTGTCATTAACAGATAGTAATACCGAGTCTTTTAAACAAAA 1039
 547 GCAAAAGAACGAAACACAAAAGTATGATGCTGATTAATCAAAAATTAACAGATAGTAT 606
 1040 GCAAAAGAACGAAACACAAAAGTATGATGCTGATTAATCAAAAATTAACAGATAGTAT 1099
 607 TTACGTAATTTGTAATACATGTAATTCCTGATTAACGCTTCTGCGGCTGGGTGTGCT 666
 1100 TTACGTAATTTGTAATACATGTAATTCCTGATTAACGCTTCTGCGGCTGGGTGTGCT 1159
 667 CAAGCACTCGGCATTTGACGAAAGAAACGATTTAAGGGAATGCTGAATGCGCGCAGAT 726
 1160 CAAGCACTCGGCATTTGACGAAAGAAACGATTTAAGGGAATGCTGAATGCGCGCAGAT 1219
 727 CCGGAGCAATGAGAAATTTCTCCGCTGATCAGTCCGACGACCTGGGCACTTTGTTAAT 786
 1220 CCGGAGCAATGAGAAATTTCTCCGCTGATCAGTCCGACGACCTGGGCACTTTGTTAAT 1279
 787 GGGTTTGGCGGAAACGACGCTTTCTTCTAATTGTAATATATGAAAAGTAAAGAAATC 846
 1280 GGGTTTGGCGGAAACGACGCTTTCTTCTAATTGTAATATATGAAAAGTAAAGAAATC 1339
 847 GGTATCCCGACCGATGATCCGATCATCATGAACTGCGCGGACGACCGGTGATGATG 906
 1340 GGTATCCCGACCGATGATCCGATCATCATGAACTGCGCGGACGACCGGTGATGATG 1399
 907 ACACGCAATTCGCAATGAGATGATGCTTATATTTGAAGAAAGTAAAGTAAATC 966
 1400 ACACGCAATTCGCAATGAGATGATGCTTATATTTGAAGAAAGTAAAGTAAATC 1459
 967 GGTGAAACAAAGAACCGATGTAAGACCTATGAAAGGCAAAATTCCTGACACAAA 1026
 1460 GGTGAAACAAAGAACCGATGTAAGACCTATGAAAGGCAAAATTCCTGACACAAA 1519
 1027 CTGCATGACCTTAGATATTAAGTCAACAGATGAAATTTAAGAAATGTTTAAAGAAATG 1086
 1520 CTGCATGACCTTAGATATTAAGTCAACAGATGAAATTTAAGAAATGTTTAAAGAAATG 1579
 1087 CACAACCGGTGATATGAGGCTGGCAATATTCATGGTGCGGACGACCTTTAATTGA 1146
 1580 CACAACCGGTGATATGAGGCTGGCAATATTCATGGTGCGGACGACCTTTAATTGA 1639
 1147 AAAATCAAGATTAACAGTAAAGCAGCTCGTAAC 1182
 1640 AAAATCAAGATTAACAGTAAAGCAGCTCGTAAC 1675
 RESULT 8
 BS292954 8189 bp DNA linear BCT 02-DEC-2004
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (bases 1 to 8189)
 Prescan,E., Moszer,I., Boursier,L., Cruz Ramon,H.C., de la
 Fuente,V., Hullo,M.F., Lelong,C., Schleich,S., Sekowska,A.,
 Song,B.H., Villani,G., Kunst,F., Danchin,A. and Glaser,P.
 The Bacillus subtilis genome from gerBC (311 degrees) to lscr (334

	/pseudo	/codon_start=1
RBS	/transl_table=11	/complement(6285..6291)
gene	/gene="ywfF"	/gene="ywfF"
RBS	6594..8112	/gene="ywfG"
CDS	6594..6601	/gene="ywfG"
	6609..7362	/note="product highly similar to metabolite transport proteins"
	/codon_start=1	/transl_table=11
	/protein_id="CAB07473.1"	/db_xref="GI:1894771"
	/db_xref="GOA:P96742"	/db_xref="UniProt/TREMBL:P96742"
	/translation="MKKSNIMLYFEFGALIGALYGYDTGVISGAILEFMKKELGNLFTEEGVLSLLNGAILSGSNAKLTDRPRKKAIAAAALLFCIGGLVALNPNTGMWLERIILGLAVGTSTIVPLVISELPKRKRGALSSLNQMTIVGILLSTVNIPADABA KRWMLGLAAPSLDLHLIGILEMPESPRMFLTNGEESFAKKI LEXLRKTKIDJSHIDI KEAEKODEGKEFLDPWVRPALIAGLGFLFOPTIGNTI IYAPRTFNVGNSA SLIGTVIGIVNVLMTVAIKIIDIKRKLLEFGNAGVISLIVLVLFENNTPA ASFTYVICGFVFIVFAVSWGPVWWMLPELFPYHVRGIVSGISTLMVHTLVSLTP YPLMEIRIGISYLEFLIYAALGIMAFLFRKRYETTKRSLEIEQDLRDKNCGGAAG KOQIVTG"	
terminator	7981..8112	
gene	/gene="ywfG"	/complement(8019..8189)
CDS	/gene="gerBC"	/complement(8019..>8189)
	/gene="gerBC"	/codon_start=1
	/transl_table=11	/product="GerBC"
	/protein_id="CAB07474.1"	/db_xref="GI:1894772"
	/db_xref="GOA:P96743"	

Query Match Best Local Similarity 89.4%; Score 1056.4; DB 1; Length 8189;
Matches 1060; Conservative 99.4%; Pred. No. 1.3e-224;

	Mismatches	Indels	Gaps
117 CCGCCGAAAATCGACTGTGCACAGGCTTCACAACCCGAAATTATAATAGAAGCCGGTTACAA	6	0	0
Db 1 CCGCGAAAAATCGACTGTGCACAGGCTTCACAACCCGAAATTATAATAGAAGCCGGTTACAA	6	0	0
Qy 177 GACGTGTGAAAAACAACAGGAACGATGCAAGAATGATTTACTGGGACACACCGGAGGA	236		
Db 61 GACTGTGAAAAACAACAGGAACGATGCAAGAATGATTTACTGGGACACACCGGAGGA	120		
Qy 237 AAAGCCGATTAAACGAAACCTCAGGGGCCGAAATATCGAGAGCAAAAAGAGTCATGAG	296		
Db 121 AAAGCCGATTAAACGAAACCTCAGGGGCCGAAATATCGAGAGCAAAAAGAGTCATGAG	180		
Qy 297 AGAAACAGTAGAAAGAGGGGCTTAACGCGATTGTCAAGTAAATGCAATGCGCTTTAACCCAGA	356		
Db 181 AGAAACAGTAGAAAGAGGGGCTTAACGCGATTGTCAAGTAAATGCAATGCGCTTTAACCCAGA	240		
Qy 357 TTATCAATCATCTTTGAGAAAGAACTTCGACGAGCCCAATATCGCGCTCATTTGTGAATGT	416		
Db 241 TTATCAATCATCTTTGAGAAAGAACTTCGACGAGCCCAATATCGCGCTCATTTGTGAATGT	300		
Qy 417 TTTAGAAACCATATGATGTGTATGAGGGCCGACGCTTGATGAAATTGCAAGCGTTTAC	476		
Db 301 TTTGAGAAACCATATGATGTGTATGAGGGCCGACGCTTGATGAAATTGCAAGCGTTTAC	360		
Qy 477 CGTACGATTCCTTTAAATGGCAATCTTGTCATTTACAGTACTGAAATATACCGAGTTCT	536		
Db 361 TGCTACGATTCCTTTAAATGGCAATCTTGTCATTTACAGTACTGAAATATACCGAGTTCT	420		

QY	537	TAAACAAAAGCAAAGAAACGAAACGAAAGATCATTTGGCTGATTAACGAAATATAC	596
Db	421	TAAACAAAAGCAAAGAAACGAAACGAAAGATCATTTGGCTGATTAACGAAATATAC	480
QY	597	AGATGAGTATTTCAGTAAATTTTGAATATACATGTATTTCCCTGATTAACGCTTCTGTGGCGCT	656
Db	481	AGATGAGTATTTCAGTAAATTTTGAATATACATGTATTTCCCTGATTAACGCTTCTGTGGCGCT	540
QY	657	GGGTGTGGCTCAAGCACTGGGCACTTTCAGAAAGAAAGCAATTTAAGGAATGCTGAATGC	716
Db	541	GGGTGTGGCTCAAGCACTGGGCACTTTCAGAAAGAAAGCAATTTAAGGAATGCTGAATGC	600
QY	717	GCCGCCAGATCCGGGGAGCAATGGAATTTCTTCGCTGATCAGTCCGAGCGAGCCTGGGCA	776
Db	601	GCCGCCAGATCCGGGGAGCAATGGAATTTCTTCGCTGATCAGTCCGAGCGAGCCTGGGCA	660
QY	777	CTTTGTTAATGGGTTTGCCGCAACGACGCTTCTTCTACTTTTGAATATATGAAGAAAGTGT	836
Db	661	CTTTGTTAATGGGTTTGCCGCAACGACGCTTCTTCTACTTTTGAATATATGAAGAAAGTGT	720
QY	837	AAAAGAAATCGGTTATCCCGACCGATGATCCGATCATCATGAATCTGCCGCGACGCG	896
Db	721	AAAAGAAATCGGTTATCCCGACCGATGATCCGATCATCATGAATCTGCCGCGACGCG	780
QY	897	TGTCGATCGGACACAGCAATTTCCGAATGACGTAATTTGCCCTTAATTTGAAGCAAGTGAAT	956
Db	781	TGTCGATCGGACACAGCAATTTCCGAATGACGTAATTTGCCCTTAATTTGAAGCAAGTGAAT	840
QY	957	GATCTTAATGGGTGAACAACAGAAACCGATCGTAAAGCCCTATGAGAAAGCAAAATTC	1016
Db	841	GATCTTAATGGGTGAACAACAGAAACCGATCGTAAAGCCCTATGAGAAAGCAAAATTC	900
QY	1017	TGCAGACAACTGCATGACCTAGATGAATGAATCAACGATGAATTTATGAATTTGTTAA	1076
Db	901	TGCAGACAACTGCATGATCTAGATGAATGAATCAACGATGAATTTATGAATTTGTTAA	960
QY	1077	GAAAGAAATGACAAACCGTGTCTATATATGGCGGTGGCAATTTTCATGATGTCGCCGAGACC	1136
Db	961	GAAAGAAATGACAAACCGTGTCTATATATGGCGGTGGCAATTTTCATGATGTCGCCGAGACC	1020
QY	1137	TTTAATTGAAAAAATTCACGAAATCAAGGTAAAGCAGCTCGTAAGC	1182
Db	1021	TTTAATTGAAAAAATTCACGAAATCAAGGTAAAGCAGCTCGTAAGC	1066

RESULT 9

AE017333_36/c

WPCOMMENT

Sequence split into 43 fragments

Fragment Name	Begin	End	LOCUS AE017333	Accession AE017333
AE017333_00	1	110000		
AE017333_01	100001	210000		
AE017333_02	200001	310000		
AE017333_03	300001	410000		
AE017333_04	400001	510000		
AE017333_05	500001	610000		
AE017333_06	600001	710000		
AE017333_07	700001	810000		
AE017333_08	800001	910000		
AE017333_09	900001	1010000		
AE017333_10	1000001	1110000		
AE017333_11	1100001	1210000		
AE017333_12	1200001	1310000		
AE017333_13	1300001	1410000		
AE017333_14	1400001	1510000		
AE017333_15	1500001	1610000		
AE017333_16	1600001	1710000		
AE017333_17	1700001	1810000		
AE017333_18	1800001	1910000		
AE017333_19	1900001	2010000		
AE017333_20	2000001	2110000		
AE017333_21	2100001	2210000		
AE017333_22	2200001	2310000		

AE017333_23 2300001 2410000
AE017333_24 2400001 2510000
AE017333_25 2500001 2610000
AE017333_26 2600001 2710000
AE017333_27 2700001 2810000
AE017333_28 2800001 2910000
AE017333_29 2900001 3010000
AE017333_30 3000001 3110000
AE017333_31 3100001 3210000
AE017333_32 3200001 3310000
AE017333_33 3300001 3410000
AE017333_34 3400001 3510000
AE017333_35 3500001 3610000
AE017333_36 3600001 3710000
AE017333_37 3700001 3810000
AE017333_38 3800001 3910000
AE017333_39 3900001 4010000
AE017333_40 4000001 4110000
AE017333_41 4100001 4210000
AE017333_42 4200001 422645
Continuation (37 of 43) of AE017333 from base 3600001 (AE017333 Bacillus licheniformis D

Query Match 65.4%; Score 772.8; DB 1; Length 110000;

Best Local Similarity 78.6%; Pred. No. 1.2e-161;

Matches 924; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

QY 7 TGGTATCATATAGCCGTGCTGCTACTGTCATCGGAATATTGAAAAAGACGA 66
DB 67582 TGGTATGCTATTAGCCTGTGTGATCTTGTGGATCGCATTTATGAAAAAGCGC 67523
QY 67 CATCAGAAAAATTGATGCTCCCTGTCGGGTGATTTAAACGGATCCGGGAAA 126
DB 67522 CACGAGAAAATTGATGCGCTGCTGTCGAGTGAATCAACGGTATACGGGAAA 67463
QY 127 TCGACTGTCAGAACGGCTGACAAACCGGAATATTATGAAACCGGTAAACACTGTTGSA 186
DB 67462 TCCACGGTGACAAAGTTAAACAGGGATATTATTCGAAAGAGGTATCAAAAACGTAGSA 67403
QY 187 AAAACAACAGAACGATGACAAAGATGTTACTGAGACACACCGGAGAAAGCCGATT 246
DB 67402 AAAACAACCGGACGACGACGAAAGATGTTATTGGACACACCGGAGAAAGCCGATC 67343
QY 247 AAACGAAACCTCAGGGGCCGAATATCGAGAGCAAAAAGATCATGAGAAAAACAGTA 306
DB 67342 AAAAGAAACCGCAGAGGGCCGAATATCGAGAGCAAGAGGATTATGAAAGAAACGCTG 67283
QY 307 GAAAGAGGGGCTAACGGCATGTGTCAGTGAATGATGCTGTTAAACCAATTTCAATC 366
DB 67282 GAAAGAGGGGCAATGCGATTGTCAGTAGTCAATGCGCTTTAAATCCTGATTACCAATC 67223
QY 367 ATCTTTGAGAAAGACTTCTGAGGGCAATATCGGCGCTATTGTAATGTTTGAAGAC 426
DB 67222 ATCTTTGAGAAAGATTGCTTCAGGCTAATATCGGCGATGTAAGAGTCTGAGGAT 67163
QY 427 CATATGATGTCATGAGGGCCGACGCTTGATGAAATTGCAAGAGCTTTACCGCTACATT 486
DB 67162 CACATGATGTCATGAGGGCCGACGCTTGATGAAATTCGAGAAAGCATTCACAGCAACATT 67103
QY 487 CTTATATATGCGCATCTTGTCAATTCAGATAGTGAATATACCGAGTTCTTTAAACAAAA 546
DB 67102 CTTATATATGACATTGTTGTTATTAAGTATGATAGTATACCGAATTCCTTTAAGCAATT 67043
QY 547 GCAAAAGAGAAACACAAAGTCAATCTGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 606
DB 67042 GCAAAAGAGAAACACAAAGTCAATCTGCTGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 66983
QY 607 TTACGTAATTTTGAATCATGTAATCCCTGATTAACGCTTCTGAGCGCTGGGTGTGCT 666
DB 66982 CTGAGAGAGTTTGAATCATGTAATCCCTGATTAATGCTGCTCTGAGCTGGGTGTGCT 66923
QY 667 CAAGCACTCGGCATTTGAGAAAGAAACAGCATTTAAAGGAATGCTGAATGCGCGCCGAT 726
DB 66922 CAAGCACTTTGAGAAAGAAACAGCATTTAAAGGAATGCTGAATGCGCGCTGAT 66863

QY 727 CCGGAGCAATGAGAAATCTTCGCTGATCATGCCAGCGCCTGGCACTTGTAT 786
DB 66862 CCGGAGCAATGAGAAATCTTCGCTGATCATGCCAGCGCCTGGCACTTGTAT 66803
QY 787 GGGTTTCCCGAAACGACGCTTCTTCACTTTGATATATGAAAGCTGTAAAAAGAAATC 846
DB 66802 GGGTTTCCCGAAACGACGCTTCTTCACTTTGATATATGAAAGCTGTAAAAAGAAATC 66743
QY 847 GGTACCCGACCGATGATCCGATCATCATGTAACCTCCGCGACGCTGTGATCGG 906
DB 66742 GGTATCTTACGATGATCCGATCATCATGTAACCTCCGCGACGCTGTGATCGG 66683
QY 907 ACACGCAATTCGCAATGACGTAATTCCTTATATGAAAGCTGATCTTATC 966
DB 66682 ACACGCAATTCGCAATGACGTAATTCCTTATATGAAAGCTGATCTTATC 66623
QY 967 GGTGAAACCAACGACCGATGTAAGGCTTATGAAAGGCAAAATTCCTGACAGCAA 1026
DB 66622 GGTGAAACCAACGACCGATGTAAGGCTTATGAAAGGCAAAATTCCTGACAGCAA 66563
QY 1027 CTGATGACCTAGATTAAGTCAACAGATTAATGAAATTTGTAAGAAAGAAATG 1086
DB 66562 CTGATGATTTTGAACCAATCAACGAAAGAAATCATGTTCACTGTAAGAAAGAAATG 66503
QY 1087 CACAACTGTCATATATGCGCTGCGCAATATTCATGTCGCGAGACCTTTAATTGAA 1146
DB 66502 GAGGCGCGCTTATTTACGAGTGGGAAATATTCAGAGCAAGCGGACCTCTCATTTGAA 66443
QY 1147 AAAATCAAGATTCAGATTAAGCACTGTTAGC 1182
DB 66442 AAAATCAAGATTCAGATTAAGCACTGTTAGC 66407

RESULT 10
CP000002_36/c
WPCOMMENT

Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000
CP000002_11	1100001	1210000
CP000002_12	1200001	1310000
CP000002_13	1300001	1410000
CP000002_14	1400001	1510000
CP000002_15	1500001	1610000
CP000002_16	1600001	1710000
CP000002_17	1700001	1810000
CP000002_18	1800001	1910000
CP000002_19	1900001	2010000
CP000002_20	2000001	2110000
CP000002_21	2100001	2210000
CP000002_22	2200001	2310000
CP000002_23	2300001	2410000
CP000002_24	2400001	2510000
CP000002_25	2500001	2610000
CP000002_26	2600001	2710000
CP000002_27	2700001	2810000
CP000002_28	2800001	2910000
CP000002_29	2900001	3010000
CP000002_30	3000001	3110000
CP000002_31	3100001	3210000
CP000002_32	3200001	3310000
CP000002_33	3300001	3410000

KPGIPARODTAUPLAEEMVAKVITAAEIKASNOENKFLLENFSEQIEVLHSHEPK
 LITIDAKAVIRITGETPIYANCILQAGVLP"
 gene complement (4777. .5658)
 /gene="fbsk"
 CDS complement (4777. .5658)
 /gene="fbsk"

Query Match 62.0%; Score 733.2; DB 1; Length 8164;
 Best Local Similarity 99.6%; Pred. No. 1.2e-152;
 Matches 735; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	7	TGGTACTCATTATGCGCTGCTGTCATCTGTCATCGGAATATTAGAAAAACGACGA	66
DB	7427	TGGTACTCATTATGCGCTGCTGTCATCTGTCATCGGAATATTAGAAAAACGACGA	7486
QY	67	CATCAGAAAAACATTTGATGCCCTCCCTGTCGGGTGAATATTAAACGATCCGCGAAAA	126
DB	7487	CATCAGAAAAACATTTGATGCCCTCCCTGTCGGGTGAATATTAAACGATCCGCGAAAA	7546
QY	127	TCGACTGTGACAAAGCTGACAAACCGGAATATTAAAGAAAGCCGTTACAGACTGTTGA	186
DB	7547	TCGACTGTGACAAAGCTGACAAACCGGAATATTAAAGAAAGCCGTTACAGACTGTTGA	7606
QY	187	AAAAACACAGAAACAGATGCAAGAAATGATTTAATGAGAAAGCCGTTACAGACTGTTGA	246
DB	7607	AAAAACACAGAAACAGATGCAAGAAATGATTTAATGAGAAAGCCGTTACAGACTGTTGA	7666
QY	247	AAACGGAACCTCAGAGGCGCAATATCGAGAGCAAAAAAGATGATGAGAGAAACAGTA	306
DB	7667	AAACGGAACCTCAGAGGCGCAATATCGAGAGCAAAAAAGATGATGAGAGAAACAGTA	7726
QY	307	GAAGAGGCGCTAACGCGATTTGTCAGTGAATGATGCTGTTAACCCAGATTATCAATC	366
DB	7727	GAAGAGGCGCTAACGCGATTTGTCAGTGAATGATGCTGTTAACCCAGATTATCAATC	7786
QY	367	ATCTTTACAGAAACCTTTGACAGCCCAATATCGCGCTGATTTGAAATTTTGAAGAC	426
DB	7787	ATCTTTACAGAAACCTTTGACAGCCCAATATCGCGCTGATTTGAAATTTTGAAGAC	7846
QY	427	CATATGATGTCATGAGGCGCGCAAGCTTTGATGAATTTGCAAGACGTTTACCGTCAAT	486
DB	7847	CATATGATGTCATGAGGCGCGCAAGCTTTGATGAATTTGCAAGACGTTTACCGTCAAT	7906
QY	487	CCTTATATGCGCATTTGTCATTAACAGATGATTAACGAGTTCTTTTAAACAAAA	546
DB	7907	CCTTATATGCGCATTTGTCATTAACAGATGATTAACGAGTTCTTTTAAACAAAA	7966
QY	547	GCAAAAGACGAAACCAAAAGTCATCTGATTAATCTCAAAATTTACAGATGAT	606
DB	7967	GCAAAAGACGAAACCAAAAGTCATCTGATTAATCTCAAAATTTACAGATGAT	8026
QY	607	TACGTAATTTTGAATACATGATTTCCCTGATTAACGCTTCTGCGCTGAGGTGAGCT	666
DB	8027	TACGTAATTTTGAATACATGATTTCCCTGATTAACGCTTCTGCGCTGAGGTGAGCT	8086
QY	667	CAAGCACTGGCATTTACAGAAAGAAACAGATTTTAAAGGAATGCTGAATGCGCGCAGAT	726
DB	8087	CAAGCACTGGCATTTACAGAAAGAAACAGATTTTAAAGGAATGCTGAATGCGCGCAGAT	8146
QY	727	CCGGAGCAATGAAATT 744	
DB	8147	CCGGAGCAATGAAATT 8164	

RESULT 12
 AX434279 819 bp DNA linear PAT 28-JUN-2002
 LOCUS
 DEFINITION Sequence 2694 from Patent WO0229113.
 ACCESSION AX434279
 VERSION AX434279.1 GI:21659087
 KEYWORDS
 SOURCE Bacillus licheniformis
 ORGANISM Bacillus licheniformis

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE
 1 Berka, R. and Clausen, I.G.
 METHODS FOR MONITORING MULTIPLE GENE EXPRESSION
 TITLE
 JOURNAL
 Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
 Location/Qualifiers
 1. 819
 /organism="Bacillus licheniformis"
 /mol_type="unassigned DNA"
 /db_xref="taxon:1402"

Query Match 46.4%; Score 548.4; DB 6; Length 819;
 Best Local Similarity 79.9%; Pred. No. 2.3e-111;
 Matches 645; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY	24	CTGTGCTGTCATCTGTCATCGGAATATTAGAAAAACGACATCGAAAAACATTTGA	83
DB	4	CTGTGCTGTCATCTGTCATCGGAATATTAGAAAAACGACATCGAAAAACATTTGA	63
QY	84	TGCCCTCCCTGTTGGGTGAATATTAAACGATCGCGGAAAAATCGACTGACAAAGCT	143
DB	64	TGCCCTCCCTGTTGGGTGAATATTAAACGATCGCGGAAAAATCGACTGACAAAGCT	123
QY	144	GACAAACCGGAATATTAAAGAAAGCCGTTACAGACTGTTGAAAAACCAAGAAACAGA	203
DB	124	AACAAACCGGAATATTAAAGAAAGCCGTTACAGACTGTTGAAAAACCAAGAAACAGA	183
QY	204	TGCAAGATGATTTAATCTGAGACACACCGAGAAAAAGCCGATTAACGAAACCTCAGG	263
DB	184	CGCAAGATGATTTAATCTGAGACACACCGAGAAAAAGCCGATTAACGAAACCTCAGG	243
QY	264	GCCGAATATCGAGAGCAAAAAAGATGATGAGAAACAGTAAAGAAAGGGGCTTAAGC	323
DB	244	GCCGAATATCGAGAGCAAAAAAGATGATGAGAAACAGTAAAGAAAGGGGCTTAAGC	303
QY	324	GATTTGTCATGATTAATCGAGCTGTTAACCCAGATTATCAATCTTTCAGAGAACT	383
DB	304	GATTTGTCATGATTAATCGAGCTGTTAACCCAGATTATCAATCTTTCAGAGAACT	363
QY	384	TCTGAGGCGCAATATCGCGCTGTCATTTGTAATGTTTAAAGCAATATGATGATGAG	443
DB	364	GCTTACGCTAATATCGCGCTGTCATTTGTAATGTTTAAAGCAATATGATGATGAG	423
QY	444	GCCGACGCTTGAATTAATGCAAGCGTTTACCGTCAATCTTCTTATATGCGCATCT	503
DB	424	ACCGACTTTGATTAATGCAAGCGTTTACCGTCAATCTTCTTATATGCGCATCT	483
QY	504	TGTCATTAACATTAATGCAATTAACCGAGTTCTTTTAAACAAAAAGCAAGAAACAC	563
DB	484	GGTTATTAACATTAATGCAATTAACCGAGTTCTTTTAAACAAAAAGCAAGAAACAC	543
QY	564	AAAAGTCATCTGTCATTAATCTCAAAATTTACAGATGATTTTACGATTTTGAATA	623
DB	544	AAAAGTCATCTGTCATTAATCTCAAAATTTACAGATGATTTTACGATTTTGAATA	603
QY	624	CATGATTTCCCTGATTAACGCTTCTGCGCTGAGGTGCTCAAGCACTCGCATTTGA	683
DB	604	CATGATTTCCCTGATTAACGCTTCTGCGCTGAGGTGCTCAAGCACTCGCATTTGA	663
QY	684	CGAAGAAACAGATTTAAAGGAATGCTGAATGCGCGCAGATTCGGGAGCAATGAAAT	743
DB	664	CGAAGAAACAGATTTAAAGGAATGCTGAATGCGCGCAGATTCGGGAGCAATGAAAT	723
QY	744	TCTTCGCTGATCAGTCGAGGAGCTGGGCACTTTGTTAATGAGTTTGGCCGAAACGA	803
DB	724	TCTTCGCTGATCAGTCGAGGAGCTGGGCACTTTGTTAATGAGTTTGGCCGAAACGA	783
QY	804	CGCTTCTTCTTGAATATATGAA 830	
DB	784	CGAGCTTCCACTTTAAACATTTGAA 810	

Or 1141 ATTGAAAAATCCAGGATA 1160
Db 1422 ATCGATATGATTATGGAAGA 1441

RESULT 14
LOCUS AE011191/c 94829 bp DNA circular BCT 14-JUN-2002
DEFINITION Bacillus anthracis str. A2012 plasmid pX02, complete sequence.
ACCESSION AE011191
VERSION AE011191.1 GI:20520280

KEYWORDS
SOURCE
ORGANISM Bacillus anthracis str. A2012
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE
AUTHORS 1 (bases 1 to 94829)
Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umayam, L., Jiang, L., Holtzapf, E., Busch, J.D., Smith, K.L., Schupp, J.M., Solomon, D., Keim, P. and Fraser, C.M.
Comparative genome sequencing for discovery of novel polymorphisms in Bacillus anthracis
Science 296 (5575), 2028-2033 (2002)

JOURNAL
PUBMED 12004073
2 (bases 1 to 94829)
Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umayam, L., Jiang, L., Holtzapf, E., Busch, J.D., Smith, K.L., Schupp, J.M., Solomon, D., Keim, P. and Fraser, C.M.
Direct Submission
Submitted (09-MAY-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source
1. 94829
/organism="Bacillus anthracis str. A2012"
/mol_type="genomic DNA"
/strain="A2012"
/db_xref="taxon:191218"
/plasmid="pX02"
/note="Florida strain"
complement(543. 1139)
/gene="BxB0001"
complement(543. 1139)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-03)"
/protein_id="AA026162.1"
/db_xref="GI:20520281"
/translation="MTYVKLFQPRGNIHVDFQFQVPEYPRADVLRLHATVLLD
EKREKGGODDVLKKEVNDLKOSVNTMKQMDVLIKLMELIAQSGNLSLEKLEO
EVASEYASSVTYKSEKFRPTVREPIEBVVOHORIQREVKTTTSVKKPLPMTIDV
QIVMKKGPHQVPTRLGTGYQSEIWEQIPEHRRKELE"
complement(1151. 1438)
/gene="BxB0002"
complement(1151. 1438)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-04)"
/protein_id="AA026163.1"
/db_xref="GI:20520282"
/translation="MEIHVANNUPYHLKEIDRCKREIGKLGRRYRWEXYNNMFGOH
FNOEYSRKKEDKDEAVSNVSTLDROSDKLQETIIDTDELVAAMTKLNEE"
complement(1652. 1768)
/gene="BxB0003"
complement(1652. 1768)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

gene
CDS
/product="hypothetical protein"
/protein_id="AA026164.1"
/db_xref="GI:20520283"
/translation="MKRKKRYKVVHPRRIQSAVRHIONAYEVLKNAKHAHTI"
complement(1746. 2594)
/gene="BxB0004"
complement(1746. 2594)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (pX02-05)"
/protein_id="AA026165.1"
/db_xref="GI:20520284"
/translation="MSTKFWNNKEGRDPAATKRAVYPLTSGPYTFMFLCISPFSG
ATYFEEKIGEVSLSTYKIAVAKENPDNQTALIDETLRADNDLSLNNKRYK
ENRYIKQDNNVAKTVRASDNVYLVSENPBEFGVSSVYKREYIHPBLQNVDL
KERSMKVYLENEKTLINRELKSKDFEERYLAFSQAALKEIEKEDSKSQAOL
KIKNEQLKEMEYQTEGEKVTKNTINSNESTINNHOKEIDVLKEDIKKERKIQOLD
EKKTITI"
complement(2591. 2866)
/gene="BxB0005"
complement(2591. 2866)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-06)"
/protein_id="AA026166.1"
/db_xref="GI:20520285"
/translation="MLSELKPNHDYAKEGYLLISLKKKGVRRKREIPIFWEDY
NPEKVEWLIYREYOPSVNGKEKYTNCLENIHQVSVVNGSRMK"
complement(2955. 3563)
/gene="BxB0006"
complement(2955. 3563)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (pX02-07)"
/protein_id="AA026167.1"
/db_xref="GI:20520286"
/translation="MNNLSKILSGALVLSFNGIMAVNTTKEPDKDKRYRVYSEK
EKQIDLVYKLEQCKKTKIVADIVAPNTDLDPRPKSVNKAATYINSGVSKCTALLD
KGERNNLLKTDKQVDIVAPNTDLDPRPKSVNKAATYINSGVSKCTALLD
IEYTIQENKQTTINSVAKITLERQSEIKVEYNYPVAK"
complement(3589. 4731)
/gene="BxB0007"
complement(3589. 4731)
/note="identified by match to PFAM protein family HMM
PF00877"
/codon_start=1
/transl_table=11
/product="NUP/p60 family protein, (pX02-08)"
/protein_id="AA026168.1"
/db_xref="GI:20520287"
/translation="MAATQETAIIDKRYKRIKWLIVRLIGSGTVVIAATLLILVS
MAIPGOSSTGPNCGISGTATVKNLPPEVVMQAMVEDECAQVPELVPLVLAIM
VSNNGISEKLPIIDINSSEOGNANNTISNPKDSIYGVNHLGAFDADAMAILINDLA
IVQTYNPGNIVVHMLAANNKTHSIOTADYSLTVVAPAGNNGTTIGVSOVAVAYN
GGRTYINGNPFYAEKVYKLSFDAGAGTSGQIPGSETPFKYMMDBVLKNGNPPVYG
GKSSSGFPCSSLTYYAKTACITTFISAATYDPTVEVDPDADQPGDLPFRGTGCG
PNHVSIVGIYIDANTMYDSNGGVGHQPTSSYWOOHVAGIRKVRP"
complement(4733. 6667)
/gene="BxB0008"
complement(4733. 6667)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (pX02-09)"
/protein_id="AA026169.1"

/db_xref="GI:20520288"
/translation="MWKLKKEQOQKRYTEGYDLDFISRVPOGQIKENEKXIAOGDCY
VACHVYLAIEDIPMLTALMNKKDITSSVDVATPANEVVKDINSITLKRMTS
ERSTDDDAHMERQNLTPARSITQGEIYKLVKRYIYDIPVLEOLERIGDIKKE
IAGNYKAOVYFKQKEBWOTLPASIDQIYLVKSGYPLPSKATGGIPFHNDLK
DPRIVYIGQSTGAFILDPFSTGTRTSFSGFI GKRGAKASTLLKQLEGLVAKDC
PIRFDKARDYVTVVOOGGKI IDLSGVQDEBEVETQINLEVEFATKITVYGVDEK
GSFIOHISKVTNMIPLATEPDDTVIOEFRGLIYAFYEMGLVPEQKSGDRYKTYGP
PNHYPILEDPEYKVISNLKEAGATPORHRELAIKOVEDMITVYGDVFNHTLLKN
ENBOIVPEIDIGISKYKSVFNCOLFETLTLTINSHALKNGBOMYLAEREKLSIEDYK
YPMVFLDECHVINSQNEFANRYLEPEREKRFAGCFEPTQSPQENLPFGSTVDY
STYKAIPELTQYKIFLNMDSVLETLKSLVLESLSEFRIPLPELKREALIVQVSTE
TYNVPFDEKQYKFKGQ"
gene
complement (6684 . 7358)
/gene="BXB0009"
complement (6684 . 7358)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pX02-10) "
/protein_id="AAM26170.1"
/db_xref="GI:20520289"
/translation="MKMKVATAEAKRVNRKQISEKSGKOSFISTIPASFKNXTDAS
IVDIAPYRLLEOKTGILVDHNNILQVLYKTTDLISNQODLKRFMNQLTSLGRVYH
EPFILSLTSETTEBOOVYKRMALRFGMSQSEVSKKEHLAYORSLAENLNR
VLWTEKNLKEIAFLIVYGNKETELIKVKNMKRYGGRFLQNMKKAKEVEKLI FKQJ
NMSEB"
gene
complement (7416 . 7730)
/gene="BXB0010"
complement (7416 . 7730)
/gene="BXB0010"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pX02-11) "
/protein_id="AAM26171.1"
/db_xref="GI:20520290"
/translation="MEMMRNPKNTKQEIKLAFPIYIIDATIALMLVLASYMPKVPYV
GFGRMFVYLFQTRGLFCIKPHNSPTNRNIFVILDMKMNKXHPLEVNTISSETK
RK"
gene
complement (7733 . 7936)
/gene="BXB0011"
complement (7733 . 7936)
/gene="BXB0011"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pX02-12) "
/protein_id="AAM26172.1"
/db_xref="GI:20520291"
/translation="NNWNIPIALMIGVAIVUSIATPFKQIFTIKKRKKKEPWVWTV
TNAEYLDREFWKSGEKQDKKEHD"
gene
complement (7938 . 8192)
/gene="BXB0012"
complement (7938 . 8192)
/gene="BXB0012"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pX02-13) "
/protein_id="AAM26173.1"
Query Match 42.9%; Score 507.2; DB 1; Length 94829;
Best Local Similarity 64.8%; Pred. No. 1.6e-102;
Matches 752; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

Db 56732 CGTTCCATCAGAAAAGGCTCAATTCTATCCCAATTGCACTAAACATTAATGCAATTCGA 56673
Qy GGAATAATGCACTGTCAGCAAGGCTGCAACACGGAATATTAATAGAGCCGGTTACAAACT 180
Db 56672 GGTAAATCTACCGTTACAAAGCTAATTAACAGTGTGTTCAAGAGAGGAATATTAACACT 56613
Qy 181 GTTGAAGAAAACACACAGAACAGATGCAGAAATGATTTTACTGGGACACACCGAGAGAAAAG 240
Db 56612 GTAGGAGAAAACAACTGGTACATCGCCCGAATGATATATTTGGTTTACTGACAGAGACAA 56553
Qy 241 CCGATTAAGCGAANAACCTCAGGGGGCGGAATATGAGAGCAAAAAGAAAGTCATGAGAA 300
Db 56552 CCGATTAGCGCCCTTAAGAGGCTCTTAATATGGTAGCAACGACGGATGTTAAAGAG 56493
Qy 301 ACACTAGAAAAGAGGGGTAAACGCGATTTGATGTAATGCAATGGCTGTTAAACCGATATAT 360
Db 56492 GCTGCTGATTTTAAAGACAGAAAGCACTTATTTGTAATGTAATGGCAGTTCAACCGATAT 56433
Qy 361 CAAATCATCTTTGAGAGAACTTTCGACGGCCAAATATCGGCGTCATTTGTAATGTTTAA 420
Db 56432 CAAATTAATCTTCCAATTAATAATGATTCAGCAAAATGTTGAGATGTTAAATGTTTAA 56373
Qy 421 GAAGACATATGATGATTCATAGGGGCCACGCTGATGTAATGCAAGAGCTTAAACCGCT 480
Db 56372 GAAGATCATATGATGATTTATAGGACCTTCACTTGAACAGAGTGAAGCTTTCACCTGCT 56313
Qy 481 ACAATTCCTTAAATGGCCATCTTGTCATTACAGATAGTAATGTAATPACCGATTTCTTAAA 540
Db 56312 ACCATTCATTAATGACATTTTATGTCATTAATGTAATGTAATPCTGGAATTACTTTAAA 56253
Qy 541 CAAAAGCAAAAGAACAAACACAAAAGTCATCTGATTAACCTGAAATTAACAT 600
Db 56252 GAGGTGAGAGAGAGAAATACAAAGATGATTTGTCGATTAATTTAGAAATTTAGAA 56193
Qy 601 GAGTATTTAGCTATTTTGAATATACATGTAATCCCTGTAATACGTTCTCGCCCTGGGT 660
Db 56192 GAAATCTTACAAAAATTTGATTAATGATGCTCTCCCAATTAATGATGCTTGCTTTAGCG 56133
Qy 661 GTGGCTCAAGCACTCGGCATTTAGCAAGAAACACAGCATTTAAGGAAATGCTGAATGCGCG 720
Db 56132 GTAGCAGAGGCTCTTGCGATTTGATGAGAAACAGCATTTCCGTGTAATGTTGAATGCTCAT 56073
Qy 721 CCAAGTCGGAGAGCAATGAGAAATCTTCCGCTGATCAGTCGACGAGACCTGGGCACTTT 780
Db 56072 CCGATCCAGAGCAATGAGAAATTAACAGTTTGTCTGACCAATTAAGCCTGCGCTTCTTC 56013
Qy 781 GTTAATGGGTTTGGCCGCAAGACGCTTCTTCTAATTGAATATATGAAACCTGTAAAA 840
Db 56012 GTTAATGGTTTTGGAGGAATGATCCCTCAACACATTAACGTATTTGGAAACGTGTGAT 55953
Qy 841 GAAATCGGTTACCCGACCGATGATCCGATCATCATCAATGAACCTGCGCGACGACCTGTCT 900
Db 55952 GATTTTGATATATGATCTAGCTCAATGTTATTAATGAATTTGCCGCCCTGACCGCGTT 55893
Qy 901 GATCGACACAGCAATTCGCAATGACGTAATGCTTATATTTGAAGCAAGTGAACGTATC 960
Db 55892 GATGTAAGCAAGCTTTGCTAGAGGATGTTTGCATATATTAAGCGGAAATGATTAT 55833
Qy 961 TTAATCGGTGAACACAGAACCAATCGTAATGAAGCCATGAAGAGCAAAATTCCTGCA 1020
Db 55832 GCGATGGAGAAAGCACTGCAACCTTAACAAGTCTTTTGAAGAAAGAAATATTCACAGC 55773
Qy 1021 GACAACTGCATGACCTAGAGTATTAAGTCAACAGATGAATATGAAATGTTAAAGAA 1080
Db 55772 CAAAGTATTTGAACCTTGAAGGCTGGTCAACAGTGAATTAATGATCTCGATCGTCA 55713
Qy 1081 AGAATGACAAACCGTGCATATATATGAGCGTCGCAATATTCATGTTGCGCAGACCTTTA 1140
Db 55712 TATTTAAATAATCGGATTTGATATGAGATGGGTAAATTAATGATGTCAGCTGACCATTA 55653
Qy 1141 ATTGAAAAATTCACGAATA 1160
Db 55652 ATCGATATGATTAAGAAAG 55633

LOCUS	DEFINITION
AE017335/c	94830 bp DNA circular BCT 09-JUL-2004
AE017335/c	Bacillus anthracis str. 'Ames Ancestor' plasmid pXO2, complete sequence.
ACCESSION	AE017335
VERSION	AE017335.3
KEYWORDS	GI:50118566
SOURCE	Bacillus anthracis str. 'Ames Ancestor'
ORGANISM	Bacillus anthracis str. 'Ames Ancestor'
REFERENCE	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
TITLE	1 (bases 1 to 94830)
JOURNAL	Ravel,J., Raeko,D.A., Shumway,M.F., Jiang,L., Cer,R.Z., Federova,N.B., Wilson,M., Stanley,S., Decker,S., Read,T.D., Salzberg,S. and Fraser,C.M.
REFERENCE	2 (bases 1 to 94830)
AUTHORS	Ravel,J., Raeko,D.A., Shumway,M.F., Jiang,L., Cer,R.Z., Federova,N.B., Salzberg,S. and Fraser,C.M.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAY-2004) Microbial Genomics, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA
REFERENCE	3 (bases 1 to 94830)
AUTHORS	Ravel,J., Raeko,D.A., Shumway,M.F., Jiang,L., Cer,R.Z., Federova,N.B., Wilson,M., Stanley,S., Decker,S., Read,T.D., Salzberg,S. and Fraser,C.M.
TITLE	Direct Submission
JOURNAL	Submitted (09-JUL-2004) Microbial Genomics, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA
REMARK	Sequence update by submitter
COMMENT	On Jul 9, 2004 this sequence version replaced gi:47552342.
FEATURES	Location/Qualifiers
SOURCE	1..94830
gene	/organism="Bacillus anthracis str. 'Ames Ancestor'"
CDS	/mol_type="genomic DNA"
gene	/strain="Ames Ancestor"; A2084"
CDS	/db_xref="taxon:261594"
gene	/plasmid="pXO2"
CDS	/note="MLVA types GT62 and A3.b"
gene	/complement(543)..1139
CDS	/locus_tag="GBAA_pXO2_0001"
gene	/complement(543)..1139
CDS	/locus_tag="GBAA_pXO2_0001"
gene	/note="Identified by Glimmer2; putative"
CDS	/codon_start=1
gene	/transl_table=1
CDS	/product="hypothetical protein, (pXO2-03)"
gene	/protein_id="AAI28931.2"
CDS	/db_xref="GI:47552343"
gene	/translation="MTYVLTQFRGNQIHVDIVDFORQYDEYPREADVIHAIVLLEEMRGKODDVSLKEVKDKLKOSVNFMKKOMDYLILNNMELLAQSSENLSLETJEEFKRSIIASVTSKSEKGPRTYRERIEENVOEHRIOREVYTTSFVKKPLPMITDDVQIVMKGRPHOVYPRLTGTGVOEIMEDQPEHRRELE"
gene	/complement(1151)..1438
CDS	/locus_tag="GBAA_pXO2_0002"
gene	/complement(1151)..1438
CDS	/locus_tag="GBAA_pXO2_0002"
gene	/note="Identified by Glimmer2; putative"
CDS	/codon_start=1
gene	/transl_table=1
CDS	/product="hypothetical protein, (pXO2-04)"
gene	/protein_id="AAI28932.2"
CDS	/db_xref="GI:47552344"
gene	/translation="MEIIRVNDVPYHLKEIDRRCKEIGKLGRYYRWYINMFECHEFNQESRNKEDGVSVSTLDRQSDKQGEVIDATHELVAAMIKINE"
gene	/complement(1652)..1768
CDS	/locus_tag="GBAA_pXO2_0003"

gene
CDS
complement (1652..1768)
/locus_tag="GBAA_px02_0003"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAT28933.2"
/db_xref="GI:47552345"
/translation="MKKKKKYKVRPRIOASVRIQNAEVLKNAKYKAHTI"
complement (1746..2594)
/locus_tag="GBAA_px02_0004"
complement (1746..2594)
/locus_tag="GBAA_px02_0004"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (px02-05)"
/protein_id="AAT28934.2"
/db_xref="GI:47552346"
/translation="MSIKFPMNKREKRPANTKRAVLTGLSGFVTFVLCISPVSG
ATYFEEKKIGEVOSLSSTYKIAVAKREKRPNDQTLRIYELPANDDSOLSMKKYK
ENKRIKKQDNNAKTKVIRASDNTIVTISENVEPEFVSSVAKPEIYIPELQDVYDL
KERMEKIVYENELINRELKKSKDFEYEELVAFSQALRKIEKMEKSSAMQOL
KIKMEQVTKMEYQTEGEKVKTKYKTIINSSTINNHQKEDIVLKDIEDIKMKKKIQLD
EKKRTI"
complement (2591..2866)
/locus_tag="GBAA_px02_0005"
complement (2591..2866)
/locus_tag="GBAA_px02_0005"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (px02-06)"
/protein_id="AAT28935.2"
/db_xref="GI:47552347"
/translation="MLSELKPNHDYAKGKYYLILSLRKKKGVAKQKFIETPTWPDY
NGEKVEMVLYREYQPSVNGKEKYTNCKLENIHQVSVNVKGERMK"
complement (2955..3563)
/locus_tag="GBAA_px02_0006"
complement (2955..3563)
/locus_tag="GBAA_px02_0006"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (px02-07)"
/protein_id="AAT28936.2"
/db_xref="GI:47552348"
/translation="MNNLSKILSGALVLSLSPNGIMAYNTTKFEKKDKKRYRVSEK
EKQIQLDLYVKLEOKDKINEGGAEKNGSTLIDLNKREVNQCVHAYLVDYSVONQ
KGERNNILKTKTDKKVYDIAPNTDDLDPRFKSHYKAIYINSBEDVSKCTALLD
LEYITLGENKQYITNSVVKITLLEKQGEILKVAEYNPVAR"
complement (3589..4731)
/locus_tag="GBAA_px02_0007"
complement (3589..4731)
/locus_tag="GBAA_px02_0007"
/note="identified by match to Pfam protein family HMW
PF00877"
/codon_start=1
/transl_table=1
/product="NLP/P60 family protein, (px02-08)"
/protein_id="AAT28937.2"
/db_xref="GI:47552349"
/translation="MAAQTETALIDKKYKVKIKRIVRLGSGTVVIAAATILLIYVS
VAIFGQSGSTPNCIGISGTAIVNKLPEVWRWQAMAEQECAAQGVBEIVYVATILTA
MNSNGISLELPIMOSSESGQAMANTTINPDOSIYGVMLHGFADAKMLGIDLLTIA
IVQTNPGRNVTMLAANNKTHSIQIADYSLVAVAPAGNRNGTITGSPVAVAYNGL
GGYRTINGENPFYAMNQVLSFPGAGTSGQIDGGSETPEVAKYKNGNXYWAG
GKSSQGFQSCSLTYWAKTIGITITPISAATQDPTFYVBDPKDAQPEDIVYFRTTGG
PNHVSVHGVIYIDANTMTDNSGSGVGQHOFISTSWOQHVAGIRVR"
complement (4733..6667)
/locus_tag="GBAA_px02_0008"
complement (4733..6667)
/locus_tag="GBAA_px02_0008"

/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (pX02-09)"
/protein_id="AAT28938.2"
/db_xref="GI:47552350"
/translation="MVKLKKEKKQKRTGYDLDFISHPQGGIKENKXIAGDGCY
VACHVYLAEDIPMLTALMNNKDTISVDVATANKKEVVDIINRITELKDRMTS
ERSTCRDPAHMLONTLPASTTQGEIYKLVLRITYIDPVLDELKRIQDKKE
IAGONYKAOVYVFKOKEMQTLFASVDQOIEYLKSGYPIPSKRTIGFIFPHNDLK
DPRGLYLGQSTGAFILDPFSTGTSTGSGFIFGKMGAKSTLLKQLEGLVAKDC
FIRGFQKADYTVVQOQGKIIDLGVQDSBEVETGMINLEVPATKITVTGAVDEK
GSFIQHSKNTMIRFLNTEFPDVIQEFRLHYAFYEMGLVPOQSGDRYKATGYR
PNHYILDFPKYISNLKLEAGATPQRHLEALIOVEDMTIVYGDWGHNTLLKNF
ENQOIVPEPDIGISKYDKSPNCOLEFLLTILMSHALKNRGMKYLREKSLIEDYK
YPMFLDECHVINSONEFANRYVLEFERNRKPAGIFPTQSGPOEHLPPGSDTVY
STMAIFELTQKIFLNMDSVLETLKVLDSLESTSEFRLLPELKREALIVQVSTSE
TYNWFDPDEKOLERFKGQ"

gene
CDS
complement (6684..7358)
/locus_tag="GBAA_PX02_0009"
complement (6684..7358)
/locus_tag="GBAA_PX02_0009"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (pX02-10)"
/protein_id="AAT28939.2"
/db_xref="GI:47552351"
/translation="MKMKVATPAKAKNRKQIIESEKNGKOSFISTIFASFKNTDAS
IVDLAPYRLBQKIGILVDHRNLDQVYKVTITDLSNMODLKRFPNQLSLCRVYH
EPFKILSTYETEGQVYKRMALRFQGMRSQVSKKEHLWYQKYSALIELNR
VLWEKNKELAFIVVYQKNETELIKVYKMKRYGQFNLQNMKAKEVERKLFLKIQ
NMNGEM"

gene
CDS
complement (7416..7730)
/locus_tag="GBAA_PX02_0010"
complement (7416..7730)
/locus_tag="GBAA_PX02_0010"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-11)"
/protein_id="AAT28940.2"
/db_xref="GI:47552352"
/translation="MEMMRNPKNTKQIKLAFYIILDAIILMLVLSYMPKVPYV
GGRIMFVLTGCTGFLPICIKPHNSPTNRNIFVILDKMKDKVHPLEVNTISSEK
RK"

gene
CDS
complement (7733..7936)
/locus_tag="GBAA_PX02_0011"
complement (7733..7936)
/locus_tag="GBAA_PX02_0011"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-12)"
/protein_id="AAT28941.2"
/db_xref="GI:47552353"
/translation="MNNMIFALMLGVAIVLSIITFPKQITFTFKRKKKEPWWTV
TNAEYLDKDFWKGEQKDKKEHD"

Query Match 42.9%; Score 507.2; DB 1; Length 94830;
Best Local Similarity 64.8%; Pred. No. 1.6e-102;
Matches 752; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

QY 1 ATGGCGTGTACTCATATAGCCGTGCTGTCATACGTCATCGGAATATTAGAAAA 60
DB 56792 ATGATCTTCATATAGTATATAGTACAGGTGTTTGTATATTATAGTATATGGAACAA 56793
QY 61 CGAGACATCAGAAAAACATTATGCGCTCCCTGTTGGGTGAATATTAAACGCGATCCGC 120
DB 56732 GGTTCCTACAGAAAAGGCTCAATTCATCCCAATTCGAGTAAACATAAATGGAATTCGA 56733
QY 121 GGAATAATGCACTGTGACAAAGCTGACAAACGGAATATTATAGAACCGCGTTACAGACT 180

DB 56672 GGTAAATCTACCGTTACAGACTAATTACAGGTGTGTGACAGAAAGCAATAATTAGACT 56613
QY 181 GTTGCAAAAACAAAGGAACGATGCAAGATGTTTACTGGGACACACCGGAGGAAAAG 240
DB 56612 GTAGGAAAACAACTGTGTACTCTGCCGAAATGATATATGTGTTTACTGACAGAGACCA 56553
QY 241 CCGATTAAACGAAAACCTCAGGGGCCCAATATCGAGAGCAAAAAGAGTCATGAGAA 300
DB 56552 CCGATTAAAGCCCGTTAAAGAAAGCTTATATGTGTGAGCAACGACAGGATAGTTAAAG 56493
QY 301 ACAGTAGAAAGGGGCTAACCGCATTTGTCAGTAATGCAATGCTGTTAACCCGATTA 360
DB 56492 GCTCTGATTTAGAGCAAGAGCACTTATTTGTGAATGATGCAAGTTCAACCGAATAT 56433
QY 361 CAATTCATCTTTGAGAAAGAACTCTGACGGCCAAATATCGGCGCATTTGTAATGTTTA 420
DB 56432 CAATTCATCTTTGAGAAATTAATATATTCAGCAAAATGTTGAGATGTTGTAATGTTTA 56373
QY 421 GAGACCATATGATGTCATGCGGCGCAGCGCTTGATGAAATTGCAAGCGTTTACCGCT 480
DB 56372 GAGATCATATGATGTTATGAGGACCTACACTTGACGAAGTGAAGCTTTCACGTCT 56313
QY 481 ACAATTCCTTATTAATGCGCATCTTGTCATTAACAGATAGTGAATATACCGATTTTAA 540
DB 56312 ACCATTCCTATTAATGACATTATGTCATTAAGAAAGTGAATCTTGATTTACTTTAA 56253
QY 541 CAAAAGCAAAAGAACGAAACACAAAAGTCATCTGCTGATACCTGAAATATACAGAT 600
DB 56252 GAGGTGACAGAGAGAGAAATACAAAGATGTTGTTGCGATTAATCTGAATTTACAA 56193
QY 601 GAGTATTTACGTAATTTTGAATATCATGTAATCCCTGATTAACGTTCTGCGCGCTGGT 660
DB 56192 GAATTCCTAGAAAATTTGATTAATCATGTCCTCCAGATATGATGATGCTTGTAGCG 56133
QY 661 GTGCTCAGCACTCGGCATGACGAAGAAACAGCATTTAAAGGAATGCTGAATGCCCG 720
DB 56132 GTACACAGAGCTCTTGAGATGATGAGAGAAACAGCATTCGCTGATGTTGAAGTCAT 56073
QY 721 CCAGATCCGGAGGAATGAGAAATCTTCCGCTGATCAGTCGAGCGAGCCGCGACCTTT 780
DB 56072 CCGATTCAGAGAGCAATGAGATTAACGTTTGCTGACCAATTTAAAGCTGCGCTTTC 56013
QY 781 GTTATGCGGTTTGGCGGAAACGACGCTTCTTCACTTTGAATATATGAAACGTGTA 840
DB 56012 GTAAATGTTTGGAGGAATGATCCCTCATCAATTAAGTATTTGGAACGTTGGA 55953
QY 841 GAAATCGGTTACCCGACCGATGATCCGATCATCATGAACCTCCGCGACACCGTGT 900
DB 55952 GATTTTGGATATAGTAATCTAGCTCAATGTAATTAATGAAATGCCCGCTGACCGGTT 55893
QY 901 GATCGACACAGCAATTCGCAAAAGAGTATGCTTATATGAAAGCAAGTGAATGATC 960
DB 55892 GATCGTACGACAGTTTGTGAGGAGTGTGTCATATATTAAGGCAAAATAGTTAT 55833
QY 961 TTAATCGGTGAACAAACGAAACCGATGCTAAACCTATGAAAGGCAAAATTCGTCA 1020
DB 55832 GCGATGAGAAACGATCGACCTTATTAACAAGTCTTTGAAAAGAGATATCCAAAG 55773
QY 1021 GACAAATGATGACCTAGAGTATAGTCAACAGATGAATATGAAATGTTAAAGAA 1080
DB 55772 CAAGGATATGGAACCTTGAAGGCTGTGCAACAGTGAATATATGTCCTGATCCGTCA 55713
QY 1081 AGAATGCAACCGTGTCAATATGCGGTGCGCAATTTATCATGTTGCCGACAGCCCTTA 1140
DB 55712 TATTTAAAAAATCGAATGATGATGAGGTGATATATTCATGTTGACGCTGACCATTA 55653
QY 1141 ATTGAAAAATTCACGAATA 1160
DB 55652 ATCGATATGATTAAGAA 55633

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 6671.32 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182
Sequence: 1 atgggcgcgttaccatcatat.....aggttaagcagctcgtgaac 1182

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 segs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_hic: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_est7: *
9: gb_gse1: *
10: gb_gse2: *
11: gb_gse3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	53.8	4.6	957	10	CM937254 TCBI6.1.H
C 2	45.6	3.9	700	9	BZ017393 oeb41a05
C 3	44.6	3.8	997	10	CNS005TE AL060767 Drosophila
C 4	44	3.7	1101	10	CNS0181N AL108773 Drosophila
C 5	43.8	3.7	807	10	CG843147 yhm3761
C 6	43.6	3.7	932	10	CG711559 TC842.1.E
C 7	43.2	3.7	969	1	A1557075 PT2.1.13
C 8	42.2	3.6	897	10	CZ545953 SRAA-ad6
C 9	42.2	3.6	930	10	CNS00023 AL060533 Drosophila
C 10	42	3.6	253	1	AU264198 AU264198
C 11	42	3.6	253	6	CF355747 lmb79f08
C 12	41.8	3.5	490	5	BU495620 PEST0ab7
C 13	41.8	3.5	556	5	BO452062 PEST0ab9
C 14	41.6	3.5	500	10	AU087992 AU087992
C 15	41.6	3.5	610	10	CM898580 RPTC12.13
C 16	41.6	3.5	760	7	CK110580 N068D08.P
C 17	41.6	3.5	786	8	DN495638 N068D08.5
C 18	41.4	3.5	1035	11	CNS04JTBG AL233821 Tetracodon
C 19	41.4	3.5	1235	8	DN705852 CL462-B01
C 20	41.2	3.5	984	10	CNS004H0 AL073646 Drosophila
C 21	41.2	3.5	987	10	CNS004I8 AL066537 Drosophila
C 22	41.2	3.5	1181	10	AG365676 Mus muscu

23	41	3.5	378	7	CO840884 LM_GLS_00
24	41	3.5	523	10	CL375123 RPTC14_47
C 25	41	3.5	774	10	AG519237 Mus muscu
C 26	41	3.5	821	2	BG622779 602647567
C 27	41	3.5	2035	10	CL082770 CH216-171
C 28	40.8	3.5	437	6	CA994326 P05-39_P8
C 29	40.8	3.5	752	2	BE876734 601488303
C 30	40.8	3.5	813	7	CO085222 GR_Ea02B
C 31	40.6	3.5	1101	10	AL063921 Drosophila
C 32	40.6	3.5	922	10	CL470900 SAIL_150
C 33	40.6	3.4	1020	6	CO050222 AGENCOURT
C 34	40.6	3.4	1101	10	CNS0039V AL063936 Drosophila
C 35	40.6	3.4	1467	10	AG280522 Mus muscu
C 36	40.4	3.4	336	1	AA014816 AG280522
C 37	40.4	3.4	669	10	CZ831413 OC_Ba021
C 38	40.2	3.4	580	5	CS85856 CS85856
C 39	40.2	3.4	676	5	C90920 C90920
C 40	40.2	3.4	743	10	CE684613 C90920
C 41	40.2	3.4	883	9	A2671124 ENTP83TF
C 42	40.2	3.4	885	9	A2685382 ENTHA38TF
C 43	40.2	3.4	890	9	A2541219 ENTGB47TR
C 44	40.2	3.4	898	9	BH139714 ENTMY38TR
C 45	40.2	3.4	913	9	BH135413 ENTMX47TF

ALIGNMENTS

RESULT 1
CM937254/c 957 bp DNA linear. GSS 20-DEC-2004
LOCUS TCBI6.1_H02_SP6 Tribolium BAC library Tribolium castaneum genomic,
DEFINITION genomic survey sequence.
ACCESSION CM937254 GI:56719867
VERSION CM937254.1
KEYWORDS GSS.
SOURCE Tribolium castaneum (red flour beetle)
ORGANISM Tribolium castaneum
REFERENCE Seward, J. and Tautz, D.
AUTHORS Tribolium castaneum BAC-ends sequencing project
TITLE Unpublished (2003)
JOURNAL
COMMENT Contact: Seward, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: seward@uni-koeln.de
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..957
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBAC3.6; Site 1: EcoRI, Site 2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN
Query Match 4.6%; Score 53.8; DB 10; Length 957;
Best Local Similarity 25.0%; Pred. No. 0.003;
Matches 142; Conservative 0; Mismatches 425; Indels 0; Gaps 0;
QY 48 AATATTAGAAAAAGCAGCAGTCAAGAAACATTGATGCCCTCCCTCGGTGCGGTGAATAT 107
Db 590 AAA 531
QY 108 TAAAGCAGTCGCCGGAATTCGACTGTGACAGGCTGACACCGAATATTAAGAGC 167

[illegible]

```

/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_id="B.oleracea002"
/notes="Vector: POTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea TO1000DPH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic

```

DNA was provided by Pablo Rabnowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

[illegible]

RESULT	3
CNS005TE/c	
LOCUS	
DEFINITION	CNS005TE 997 bp DNA linear GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence TET3
KEYWORDS	BACR1XK22 of RPII-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
SOURCE	AJ060767
ORGANISM	GI:4943573
	GSS.
	Drosophila melanogaster (fruit fly)
	Drosophila melanogaster

REFERENCE
1 (baes 1 to 997)

COMMENTS	AUTHORS TITLE JOURNAL
<p>Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)</p> <p>Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oosagawa and Aaron Mammosses at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw; sp. The same strain used for the library p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.</p>	

```
FEATURES
source
    location/Qualifiers
    1..997
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
    /db_xref="taxon:7227"
    /clone="BACR12K23"
    /clone_1lb="RCC1.98"
    /note="end : TcR3"
```

[illegible]

[illegible]

Y	278	ACCAAAAGAGTCAATGAGAGAAACAGTGTGAAGAGGGGCTAACGGGATGTGATGCA	335
LOCUS	CG843147		
DEFINITION	3', genomic survey sequence.		
ACCESSION	CG843147		
VERSION	CG843147.1		
KEYWORDS	GI:51788723		
SOURCE	Bos taurus (cow)		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.		
AUTHORS	Choi, I.		
TITLE	Genomic sequences from Korean Cattle (Hanwoo) blood		
JOURNAL	Unpublished (2003)		
COMMENT	Contact: Inho Choi Molecular Biology Yeungnam University 214-1, Dae-dong Gyeongsan, Korea 712-749 Tel: 82 53 810 2933 Fax: 82 53 816 3637 Email: inhochoi@ynmail.ac.kr Seq primer: RP2 Reverse Class: BAC ends.		
FEATURES	location/Qualifiers		
source	1..807		
	/organism="Bos taurus"		
	/mol_type="genomic DNA"		
	/strain="Korean Cattle (Hanwoo)"		
	/db_xref="taxon:9913"		
	/clone="HW-YUBAC2-188-2-C04-RP2"		
	/sex="Male"		
	/tissue_type="Blood"		
	/cell_type="Leucocyte"		
	/dev_stage="Adult"		
	/clone_lib="HW-YUBAC"		
	/note="Vector: pindigobAC-5"		
ORIGIN			
Query Match	3.7%; Score 43.8; DB 10; Length 807;		
Best Local Similarity	62.2%; Pred. No. 1.6;		
Matches	69; Conservative 0; Mismatches 42; Indels 0; Gaps 0;		
Y	160	ATGAGAGCCGGTTACAGACTGTTGGAAAAACACAGAACAGATGCAAGAATGATTAC	219
Db	690	AAATTAGATGTTTACACACTGTGGGAGACTGTGGAACTTTTCAAGGCTAGTCTC	749
Y	220	TGGGACACACGGAGAAAGCCGATTAAACGAAACCTCAGGGCCGAAT	270
Db	750	CTGGAATAGATTAAGAAAGGCGTGTTCATGGAGCCACAGAGACTTAAT	800
RESULT 5			
LOCUS	CG771559	932 bp	DNA linear GSS 01-SEP-2004
DEFINITION	Tc842.1_E10.SP6 Tribolium BAC library Tribolium castaneum genomic, genomic survey sequence.		
ACCESSION	CG771559		
VERSION	CG771559.1		
KEYWORDS	GI:38024738		
SOURCE	Tribolium castaneum (red flour beetle)		
ORGANISM	Tribolium castaneum		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenetbrionidae; Tribolium.		

REFERENCE 1 (bases 1 to 932)
AUTHORS Savard, J. and Tautz, D.
TITLE Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savard, J.
Abteilung für Evolutionsgenetik, AG Tautz
Institut für Genetik, Universität zu Köln
Weyertal 121, 50931 Köln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.

FEATURES
source location/Qualifiers
1..932
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="taxon:7070"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN
Query Match 3.7%; Score 43.6; DB 10; Length 932;
Best Local Similarity 56.2%; Pred. No. 1.9;
Matches 82; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 943 GAAGCAAGTGAATCTTATCGGTGAACACAGAACCGATGTAAGCTATGAA 1002
DB 359 GCAGAACTGAATATTATGAATAATGCTCAAAACCGAATCGAAGTGAAGATAGAA 418
QY 1003 GAAGCAAAATCTTCGACGACAACTGATGATGATGATGATGATGATGATGATGAT 1062
DB 419 GAGGAAAAAGTTAAAGAGTCAAGAGAAAAAGTCAAGAGTCAAGAGAGAACT 478
QY 1063 ATGGAATGTTAAGAAAAAGATGCA 1088
DB 479 AAGGACTTGAACAGAAAAAGTTCA 504

RESULT 7
A1557075 969 bp mRNA linear EST 09-AUG-1999
LOCUS PT2.1.13 C03.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION A1557075
VERSION A1557075.1 GI:4489438
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 969)
Huang, G.M., Ng, W.L., Farkas, J., He, L., Liang, H.A., Gordon, D., Yu, J.
and Hood, L.
Prostate cancer expression profiling by cDNA sequencing analysis
Genomics 59 (2), 178-186 (1999)
10409429
Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
Location/Qualifiers
1..969
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional

ORIGIN
Query Match 3.7%; Score 43.2; DB 1; Length 969;
Best Local Similarity 43.0%; Pred. No. 2.4;
Matches 114; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 48 AATATTGAAAAACGACGACATCGAAAAACATTGATGCCCTCCCTGTTGGGGGAAATAT 107
DB 531 AAGNANTNA 590
QY 108 TAAAGCATCCGCGGAAATCGACTGTGACAGGCTGACACCGAATATTATAGAAC 167
DB 591 AAAAAAAAAAGAAANTTAACCCCAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAGN 650
QY 168 CGGTTACAGACTGTTGAAAAACACAGACAGATGCAAGATGATTTACTGGGACAC 227
DB 651 GGAANNCAAAACAAAGAACAGAAAAAAGAAAAAAGNAAAAATAGAAAAAA 710
QY 228 ACCGAGAGAAAGCCGATTAAACGAAACCTCAGGGCGCAATATCGAGAGCAAAAAA 287
DB 711 AAATTGAAAAAGAAAAAGAAAGAAAGAAAGGCAAGGAGAAANNAATACAGAAAGNAA 770
QY 288 AGTCATGAGAGAAACAGTAGAAGA 312
DB 771 AAATTAAAAAANNAANNAANNAANNA 795

RESULT 8
C2545953 897 bp DNA linear GSS 13-MAY-2005
LOCUS SRRA-aad65a06.b1 Strongyloides ratci whole genome shotgun library
(SRAGSS 004) Strongyloides ratci genomic, genomic survey sequence.
ACCESSION C2545953
VERSION C2545953.1 GI:64679422
KEYWORDS GSS.
SOURCE Strongyloides ratci
ORGANISM Strongyloides ratci
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Strongyloides.
1 (bases 1 to 897)
Mitreva, M., McCarter, J.P., Thompson, F., Viney, M., Pape, D.,
Ritter, E., Martin, J., Wylie, T., Dante, M., Waterston, R.H.,
Clifton, S.W. and Wilson, R.
Genome Survey sequences from the rat parasitic nematode
Strongyloides ratci
Unpublished (2005)
Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@wustl.edu
Genomic DNA was provided by Fiona Thompson
(F.Thompson@bristol.ac.uk) and Mark Viney
(Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol,
UK.
Class: shotgun.
Location/Qualifiers
1..897
/organism="Strongyloides ratci"
/mol_type="genomic DNA"
/strain="isofemale line ED321 heterogonic"
/db_xref="taxon:34506"
/dev_stage="infective larval stage (L3)"
/lab_host="GSI0"
/clone_lib="Strongyloides ratci whole genome shotgun
library (SRAGSS 004)"
/note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
Strongyloides ratci genomic DNA was randomly sheared,

ACCESSION CF355747
VERSION CF355747.1 GI:34022211
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 263)
AUTHORS Tidwell,R., Clifton,S., Marra,M., Hillier,L., Page,D., Martin,J., Wylie,T., Theising,B., Bowers,Y., Gibbons,M., Ritzer,E., Bennet,J., Ronko,I., Teagareishvili,R., Belaygorod,L., Grow,A., Maguire,L., Wareson,R. and Wilson,R.
TITLE Maehu Stem cell EST Project
JOURNAL Unpublished (2002)
COMMENT Contact: Jeff Gordon and Mike Lovett
Maehu, Human Genetics Division
Washington University School of Medicine
1st strand of cDNA was synthesized with reverse transcriptase and oligo(dT) beads, then cDNA was amplified by PCR using modified SMART primers. The final cDNA was cloned in PAMPI vector in annealing reaction with Uracil DNA Glycosylase (UDG). Library constructed by Y.Korshunova and M. Lovett. Library materials provided by Mills JC & Gordon JI.
FEATURES
source
1..263
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="Gastric Epithelial Progenitor"
/dev_stage="adult"
/lab_host="DH5alpha"
/clone_lib="Gastric Epithelial Progenitor"
/note="Vector: PAMPI. This library was created from laser-captured isothermal cells from tox176 transgenic mice. 1st strand of cDNA was synthesized with reverse transcriptase and oligo(dT) beads, then cDNA was amplified by PCR using modified SMART primers. The final cDNA was cloned in PAMPI vector in annealing reaction with Uracil DNA Glycosylase (UDG). Library constructed by Y.Korshunova and M. Lovett. Library materials provided by Mills JC & Gordon JI."

ACCESSION BU495620
VERSION BU495620.1 GI:22791814
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 490)
AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Page,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritzer,E., Bennet,J., Jentle,E., Ronko,I., Teagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Madkins,J., Kennedy,S., Levinso,D., Wareson,R., Wilson,R. and Sibley,D.
TITLE Maehu Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
Maehu Plasmodium EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (lsibley@orcim.wustl.edu), Washington University
Trace considered overall poor quality
Seq primer: -400P from Gibco
High quality sequence stop: 1.
location/Qualifiers
1..490
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/db_xref="taxon:36329"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2: XhoI. Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin (0.1%)-lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAasist helper phage (Stratagene), the phagemids were preplated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

Db 159 AATATGATAAATATTCATGATATATTTGTTTATTAGTTTATATTAATTAATGAGAA 100
QY 518 GTGAAATATACCGAGTTCTTTAAACAAAGCAAGCAACAAATGTCATG 577
Db 99 ATGTATTAACCAACCCCAATAAAAATTAATAAAAAATTAATTAATTT 40
QY 578 CTGATAACTCAAAATTCAGATGAGTATTTA 609
Db 39 ATATATATAATAATTAATTAATTAATTA 8

RESULT 15

CM898580 610 bp DNA linear GSS 08-DEC-2004
CM898580
LOCUS RPCI42_133L22.TJ RPCI-42 Bos taurus genomic clone RPCI42_133L22,

DEFINITION genomic survey sequence.
ACCESSION CM898580
VERSION CM898580.1 GI:56440850

KEYWORDS GSS.
SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 610)
Larkin,D.M., Donthu,K., LeDuc,R., Ryan,K., Liu,L., de Jong,P.J. and

REFERENCE

AUTHORS

TITLE End sequencing of Holstein BAC library RPCI-42
JOURNAL Unpublished (2004)
COMMENT Other GSSs: RPCI42_133L22.TV

CONTACT: Harris Lewin

Department of Animal Sciences
University of Illinois at Urbana Champaign

1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 333 5998
Fax: 217 244 5617

Email: h-lewin@uiuc.edu

Clones are derived from the bovine BAC library RPCI-42
(http://bacpac.chori.org/mbvovine42.htm). For BAC library

availability, please contact Pieter de Jong
(pdejong@gmail.com). Clones may be purchased from BACPAC

Resources (http://www.chori.org/bacpac/ordering_information.htm).
Funding was provided in part by grant no. AG2004-34480-14417 from

USDA-CGRS (Livestock Genome Sequencing Initiative) and
AGS8-5438-2-313 from USDA-ARS

Plate: 133 row: L column: 22
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers

1..610
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="RPCI42_133L22"
/sex="male"
/cell_type="blood"
/clone_id="RPCI-42"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
RPCI-42 Bovine BAC library (Male) produced by Pieter de
Jong"

ORIGIN

Query Match 3.5%; Score 41.6; DB 10; Length 610;

Best Local Similarity 54.6%; Pred. No. 6;
Matches 83; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 475 ACCGTTAATTCCTTATATGCGCATCTTGCTTACAGATGATATATACGAGTTC 534
Db 358 ACCCATCTTCCCTTCTGTTGCTTACATTAATTCAGGCTTAAATATATACACTTT 417
QY 535 TTTAAACAAAAGCAAAAGCAAAAGCAAAAGTCATGCTGATTAATCAAAATTT 594

Db 418 TTTAAATCAAAAGAGCAAAATGATATCAAAAGGAGAAATATTTTGAAGTGAACACT 477
QY 595 ACAGATGATATTTAGCTAAATTTTGAATACAT 626
Db 478 AATGACACTTTGCTTACCTTAATGATACAT 509

Search completed: February 27, 2006, 11:09:21
Job time : 6675.32 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time 746.204 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182

Sequence: 1 atggctggttaccatcatat.....agtaagcagctcgtaagc 1182

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2002as:*
- 6: Geneseq2003as:*
- 7: Geneseq2004as:*
- 8: Geneseq2005as:*
- 9: Geneseq2006as:*
- 10: Geneseq2007as:*
- 11: Geneseq2008as:*
- 12: Geneseq2009as:*
- 13: Geneseq2010as:*
- 14: Geneseq2011as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1182	100.0	1182	10	ADA44791 Bacillus
2	1182	100.0	1182	12	AD007641 B. subtilis
3	1182	100.0	1182	13	ADV65732 B. subtilis
4	1182	100.0	6536	10	ADA44796
5	1171.2	99.1	1182	4	AAH82255
6	1171.2	99.1	3045	4	AAH82254
7	548.4	46.4	819	6	ABK75403
8	338.6	28.6	1164	6	ABN91998
9	338.6	28.6	1164	13	AD501519
10	338.6	28.6	3240	4	AAH54379
11	313.2	26.5	1212	9	ADB86113
12	313.2	26.5	110000	9	ADB12064
13	294.2	24.9	560	13	ADR48536
14	246.2	20.8	861	4	AAH53464
15	181	15.3	309	6	ABK75876
16	52.6	4.5	2000	8	ADA71938
17	44.6	3.8	2000	11	ACU37108
18	44.4	3.8	1338	8	ACA42892
19	42.6	3.6	600	13	ACN56103

ALIGNMENTS

C 20	42.2	3.6	434	5	ADL44027	ADL44027 Human ova
C 21	42	3.6	12507	6	ABL32299	ABL32299 Human imm
C 22	41	3.5	6000	6	ABL34305	ABL34305 Human imm
C 23	40.8	3.5	2000	8	ADA71938	ADA71938 Rice gene
C 24	40.2	3.4	112	13	ADR48527	ADR48527 capB targ
C 25	40	3.4	7312	6	ABL33815	ABL33815 Human imm
C 26	40	3.4	7312	6	ABL70402	ABL70402 Chemical11
C 27	40	3.4	7312	6	AAH54350	AAH54350 Human gen
C 28	39.8	3.4	11944	6	ABL34186	ABL34186 Human imm
C 29	39.6	3.4	1047	6	ABN93213	ABN93213 Staphyloc
C 30	39.6	3.4	1047	13	AD504156	AD504156 Staphyloc
C 31	39.6	3.4	1546	10	ADC87518	ADC87518 Human GPC
C 32	39.6	3.4	4642	4	AAH54726	AAH54726 S. epider
C 33	39.2	3.3	13133	6	ABK31230	ABK31230 Signal tr
C 34	39.2	3.3	3095	2	AAO03875	AAO03875 Sequence
C 35	38.8	3.3	7057	4	AAH46601	AAH46601 Tumour su
C 36	38.8	3.3	7057	6	ABL33848	ABL33848 Human imm
C 37	38.8	3.3	7057	6	ABL34594	ABL34594 Human met
C 38	38.8	3.3	7057	6	ABL70407	ABL70407 Chemical11
C 39	38.8	3.3	7057	6	AAH54354	AAH54354 Human gen
C 40	38.8	3.3	7057	7	ADS99855	ADS99855 Bistulphit
C 41	38.6	3.3	1008	8	ACA46836	ACA46836 Prokaryot
C 42	38.6	3.3	1011	4	AAH53908	AAH53908 S. epider
C 43	38.6	3.3	5311	6	ABL33019	ABL33019 Human imm
C 44	38.6	3.3	103929	3	AAF22287	AAF22287 BAC conta
C 45	38.4	3.2	511	13	ACN45990	ACN45990 Cotton pr

RESULT 1
ADA44791
ID ADA44791 standard; DNA; 1182 BP.

ADA44791;

20-NOV-2003 (first entry)

Bacillus subtilis poly-gamma-glutamate synthetase complex psbB gene.

Cell surface; expression vector; microbial;
KW poly-gamma-glutamate synthetase; psbB; surface expression;
KW Gram-positive bacterium; Gram-negative bacterium; enzyme; antigen;
KW antibody; attachment protein; adsorption protein; vaccine; psbB; gene;
ds.

Bacillus subtilis; variety chungkookjang.

WO2003014360-A1.

20-FEB-2003.

09-AUG-2002; 2002WO-KR001522.

10-AUG-2001; 2001KR-0048373.

(BIOL-) BIOLEADERS CORP.

(MDMP-) MD LAB CO LTD.

Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashuchi M;

WPI; 2003-256589/25.

New expression vector containing gene(s) that encode a poly-gamma-glutamate synthetase complex, useful for producing proteins (e.g. vaccines or enzymes) on the microbial surface of Gram-positive and/or Gram-negative bacteria.

Claim 3; Page 109-110; 122pp; English.

The invention relates to a vector for expression of a target protein on a microbial cell surface. The vector of the invention comprises either one

or more than two genes of the *Bacillus subtilis* poly- γ -glutamate synthetase complex (psbBCA) to facilitate microbial surface expression of the target protein. The psbBCA gene complex comprises the psbB, psbB and psbA genes and is normally expressed in the outer membrane of *Bacillus subtilis*. The vector can be transformed into either Gram-positive or Gram-negative bacteria (e.g., *Bacillus coli*), and can be used for the surface expression of various proteins of interest such as enzymes, antigens, antibodies, attachment proteins or adsorption proteins. CC proteins recombinantly produced using the vector of the invention can be used as, for example, vaccines or enzymes. The present sequence CC represents the *Bacillus subtilis* psbBCA complex gene psbB, which is specifically claimed for use in the vector of the invention.

Sequence 1182 BP; 396 A; 237 C; 274 G; 275 T; 0 U; 0 Other;

Query Match 100.0%; Score 1182; DB 10; Length 1182;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 ATGGGCTGTACTACATATTATAGCTGTGTCATCTGTCATCGAATATTGAAAA 60
Db 1 ATGGGCTGTACTACATATTATAGCTGTGTCATCTGTCATCGAATATTGAAAA 60
Oy 61 CGACGACATCAGAAAAACATTGATGCCCTCCCTGTTCCGGTGAATATTACCGCATCCG 120
Db 61 CGACGACATCAGAAAAACATTGATGCCCTCCCTGTTCCGGTGAATATTACCGCATCCG 120
Oy 121 GGAATATGATCTGTACAGAGCTGTACACCGGAATATTATGAAACCGGTTACAGACT 180
Db 121 GGAATATGATCTGTACAGAGCTGTACACCGGAATATTATGAAACCGGTTACAGACT 180
Oy 181 GTTGGAAAAACACAGAAACAGATGCAAGATGATTTACTGCGAGACACCGGAGGAAAG 240
Db 181 GTTGGAAAAACACAGAAACAGATGCAAGATGATTTACTGCGAGACACCGGAGGAAAG 240
Oy 241 CCGATTTAAACGAAAACCTCAGGGGCCGAATATCGAGAGCAAAAAGAGTCATGAGAGA 300
Db 241 CCGATTTAAACGAAAACCTCAGGGGCCGAATATCGAGAGCAAAAAGAGTCATGAGAGA 300
Oy 301 ACAGTAGAAAAAGGGGCTTACGCCATTTGTCACTGATGATGCGCTTTAACCCAGATTAT 360
Db 301 ACAGTAGAAAAAGGGGCTTACGCCATTTGTCACTGATGATGCGCTTTAACCCAGATTAT 360
Oy 361 CAAATCATCTTTCAGAGAGAACTTCTGAGGCAATATCGGGTCATTTGTAATTTT 420
Db 361 CAAATCATCTTTCAGAGAGAACTTCTGAGGCAATATCGGGTCATTTGTAATTTT 420
Oy 421 GAAGACATATGATGTATGGGGCCGACGCTTGAATGAATTTGCAAGCGTTTACCGCT 480
Db 421 GAAGACATATGATGTATGGGGCCGACGCTTGAATGAATTTGCAAGCGTTTACCGCT 480
Oy 481 ACAATTCCTTATATAGCCATTTGTCTATACAGATAGTGAATATCCGAGTTCTTTTAA 540
Db 481 ACAATTCCTTATATAGCCATTTGTCTATACAGATAGTGAATATCCGAGTTCTTTTAA 540
Oy 541 CAAAAAGCAAAAGACGAAACGAAAGTCATCTGTCATTAATCACTAAAAATTACAGAT 600
Db 541 CAAAAAGCAAAAGACGAAACGAAAGTCATCTGTCATTAATCACTAAAAATTACAGAT 600
Oy 601 GAGTATTTACGTAATTTTGAATACATGATATCCCTGATTAACGCTTCTGCGCTGGGT 660
Db 601 GAGTATTTACGTAATTTTGAATACATGATATCCCTGATTAACGCTTCTGCGCTGGGT 660
Oy 661 GTGGCTCAGCACTGGCATTTGACAGAAACAGATTTAAGGGAATGCTGAATCGCGG 720
Db 661 GTGGCTCAGCACTGGCATTTGACAGAAACAGATTTAAGGGAATGCTGAATCGCGG 720
Oy 721 CCAGATCCGGGGAGCAATGAGATTTCTCCGCTGATCAGTCGAGGAGCGCTGGGACATT 780
Db 721 CCAGATCCGGGGAGCAATGAGATTTCTCCGCTGATCAGTCGAGGAGCGCTGGGACATT 780
Oy 781 GTTAATGCGTTTCCGCAACGACGCTTCTTACTTTGAATATATGAAACGTGTAAAA 840
Db 781 GTTAATGCGTTTCCGCAACGACGCTTCTTACTTTGAATATATGAAACGTGTAAAA 840

```

```

Db 781 GTTAATGCGTTTCCGCAACGACGCTTCTTACTTTGAATATATGAAACGTGTAAAA 840
Oy 841 GAAATCGTTATCCGACCGATGATCCGATCATCATCATGAACTGCCGCGAGACCGTGT 900
Db 841 GAAATCGTTATCCGACCGATGATCCGATCATCATCATGAACTGCCGCGAGACCGTGT 900
Oy 901 GATCGACACAGCAATTCGCAATGACGATATGCTTATATTTGAAGCAATGAACTGATC 960
Db 901 GATCGACACAGCAATTCGCAATGACGATATGCTTATATTTGAAGCAATGAACTGATC 960
Oy 961 TTAATCGGTGAACACAGAACCGATGTAAAGCCCTATGAAAGGCAAAATTCCTGCA 1020
Db 961 TTAATCGGTGAACACAGAACCGATGTAAAGCCCTATGAAAGGCAAAATTCCTGCA 1020
Oy 1021 GACAACTGTCATGCTTAGATATTAAGTCAACAGATGAATTAATGAAATTTTAAAGAA 1080
Db 1021 GACAACTGTCATGCTTAGATATTAAGTCAACAGATGAATTAATGAAATTTTAAAGAA 1080
Oy 1081 AGAATGCAACAACCGTGTATATATGCGGTGCGCAATATGATGTCGCGAGACCTTTA 1140
Db 1081 AGAATGCAACAACCGTGTATATATGCGGTGCGCAATATGATGTCGCGAGACCTTTA 1140
Oy 1141 ATTGAAAAATCCACGAATACAGGTAAAGCAGCTGTAAGC 1182
Db 1141 ATTGAAAAATCCACGAATACAGGTAAAGCAGCTGTAAGC 1182

```

RESULT 2
AD007641 standard; DNA, 1182 BP.
AD007641;
15-JUL-2004 (first entry)
B subtilis poly-X-glutamate synthetase complex coding sequence psbB.
ds; gene; enzyme; vaccine; cytostratic; psbB; poly-X-glutamate synthetase;
human papilloma virus.
Bacillus subtilis.
WO2004035795-A1.
29-APR-2004.
17-OCT-2003; 2003WO-KR002163.
17-OCT-2002; 2002KR-00063378.
(BIOL-) BIOLEADERS CORP.
(KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
Sung M, Poo H, Lee J, Jung C, Hong S, Kim C, Park S, Pyo H;
WPI; 2004-348463/32.
New vector containing psb A-C genes encoding poly- γ -glutamate synthetase complex and an antigen protein gene of human papilloma virus, useful in preparing vaccine for treating or preventing mucosal tumor, e.g. cervical cancer.
Disclosure; Page 60-61; 69pp; English.
The present invention relates to a vector for preparing a vaccine which contains one or more than two genes, i.e. psb A-C encoding poly-X-glutamate synthetase complex and an antigen protein gene of human papilloma virus. The vector and microbes transformed with it are useful in preparing vaccines for treating or preventing mucosal tumor, e.g. cervical cancer. The present sequence is a *Bacillus subtilis* poly-X-glutamate synthetase complex coding sequence.
Sequence 1182 BP; 396 A; 237 C; 274 G; 275 T; 0 U; 0 Other;


```
QY 61 CGACGACATCAGAAAAAATTGATGCCCTCCCTGTTCCGGGTGAATATTAAAGCGATCCGC 120
DB 61 CGACGACATCAGAAAAAATTGATGCCCTCCCTGTTCCGGGTGAATATTAAAGCGATCCGC 120
QY 121 GGAATAATGACTGTGACAAAGCTGACAAACCGAATATTAAATGAAGCCGTTTACAAGCT 180
DB 121 GGAATAATGACTGTGACAAAGCTGACAAACCGAATATTAAATGAAGCCGTTTACAAGCT 180
QY 181 GTTGGAAAAACAACGAGAAAGATGACAAAGATTTTCTGAGGACACACCGGAGGAAAAAG 240
DB 181 GTTGGAAAAACAACGAGAAAGATGACAAAGATTTTCTGAGGACACACCGGAGGAAAAAG 240
QY 241 CCGATTAAACGGAAACCTCAGGGGCGAATATCGAGAGCAAAAAAGAAAGTATGAGAGAA 300
DB 241 CCGATTAAACGGAAACCTCAGGGGCGAATATCGAGAGCAAAAAAGAAAGTATGAGAGAA 300
QY 301 ACAGTAGAAGAGGGGCTTAACCGCATTTGTCAAGTGAATGATGCTGTTAAACCCAGATTAT 360
DB 301 ACAGTAGAAGAGGGGCTTAACCGCATTTGTCAAGTGAATGATGCTGTTAAACCCAGATTAT 360
QY 361 CAATTCATCTTTTCAAGAACTTCTGAGGCGCAATATCGCGCATTTGTAATGTTTAA 420
DB 361 CAATTCATCTTTTCAAGAACTTCTGAGGCGCAATATCGCGCATTTGTAATGTTTAA 420
QY 421 GAAGACCATATGATGATGAGGGCGAGCGCTTGATGAATTTGCAAGCGTTTAAACCGCT 480
DB 421 GAAGACCATATGATGATGAGGGCGAGCGCTTGATGAATTTGCAAGCGTTTAAACCGCT 480
QY 481 ACAATTCCTTAATATGAGCATCTTGTCTTAACAGATAGTGAATATCCGAGTTCTTTAA 540
DB 481 ACAATTCCTTAATATGAGCATCTTGTCTTAACAGATAGTGAATATCCGAGTTCTTTAA 540
QY 541 CAAAAGCAAAAAGAAAGCAAAACCAAAAGTCAATGCTGATGATCAAAAATTAACAGAT 600
DB 541 CAAAAGCAAAAAGAAAGCAAAACCAAAAGTCAATGCTGATGATCAAAAATTAACAGAT 600
QY 601 GAGATTATTTACGTAATTTTGAATACATGATTTCCCTGATTAACGCTTCTGCGCGTGGGT 660
DB 601 GAGATTATTTACGTAATTTTGAATACATGATTTCCCTGATTAACGCTTCTGCGCGTGGGT 660
QY 661 GTGGCTCAGACACTCGGCATTTGACGAAAGAAACAGCATTTAAAGGAGATCTAAATGCGCG 720
DB 661 GTGGCTCAGACACTCGGCATTTGACGAAAGAAACAGCATTTAAAGGAGATCTAAATGCGCG 720
QY 721 CCAATTCGGGAGCAATGAAATTTCTCGCATGATCCGAGGAGCACTGCGGACCTT 780
DB 721 CCAATTCGGGAGCAATGAAATTTCTCGCATGATCCGAGGAGCACTGCGGACCTT 780
QY 781 GTTAATGGGTTTGGCGCAAAAGCAAGCTTCTTCTACTTTGAATATATGAAGCTGTAAAA 840
DB 781 GTTAATGGGTTTGGCGCAAAAGCAAGCTTCTTCTACTTTGAATATATGAAGCTGTAAAA 840
QY 841 GAAATTCGGTTATGCGCAACGATGATCCGATCATCAATGAACTGCGCGCAGACCGTGT 900
DB 841 GAAATTCGGTTATGCGCAACGATGATCCGATCATCAATGAACTGCGCGCAGACCGTGT 900
QY 901 GATGGGACACAGCAATTTGCAAAATGACGATTTGCTTTATTTGAAGCAAGTGAATC 960
DB 901 GATGGGACACAGCAATTTGCAAAATGACGATTTGCTTTATTTGAAGCAAGTGAATC 960
QY 961 TTAATTCGGTGAACAAACAGAACCGATCGTAAAGGCTTAAGAAAGCAAAATTCCTCA 1020
DB 961 TTAATTCGGTGAACAAACAGAACCGATCGTAAAGGCTTAAGAAAGCAAAATTCCTCA 1020
QY 1021 GACAAATCGATGACCTAGAGTATAGTCAACAGATGAATTTATGAAATTTGTTAAAGAA 1080
DB 1021 GACAAATCGATGACCTAGAGTATAGTCAACAGATGAATTTATGAAATTTGTTAAAGAA 1080
QY 1081 AGAATGCAACACCGTGTATATATGCGCGCAATATTTCAATGATGCGCGCAGAGCTTTA 1140
DB 1081 AGAATGCAACACCGTGTATATATGCGCGCAATATTTCAATGATGCGCGCAGAGCTTTA 1140
```

```
QY 1141 ATTGAAAAAATCCACGAATACAAAGTAAACGAGCTGCTAAGC 1182
DB 1141 ATTGAAAAAATCCACGAATACAAAGTAAACGAGCTGCTAAGC 1182

RESULT 4
ADA44796
ID ADA44796 standard; DNA; 6536 BP.
AC
XX ADA44796;
XX
XX
DT 20-NOV-2003 (first entry)
XX
XX
DE Microbial cell surface expression vector pGNBCA, SEQ ID NO:6.
XX
XX
KW Cell surface; expression vector; microbial;
KW poly-gamma-glutamate synthetase; Bacillus subtilis pgsBCA complex;
KW surface expression; Gram-positive bacterium; Gram-negative bacterium;
KW enzyme; antigen; antibody; attachment protein; adsorption protein;
KW vaccine; pGNBCA; cyclic; circular; ds.
XX
OS Synthetic.
OS Bacillus subtilis; variety chungkookjang.
XX
XX WO2003014360-A1.
XX
XX 20-FEB-2003.
XX
XX 09-AUG-2002; 2002MO-KR001522.
XX
XX 10-AUG-2001; 2001KR-00048373.
XX
XX (BIOL-) BIOLEADERS CORP.
XX
XX (MDMD-) MD LAB CO LTD.
XX
XX Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashuchi M;
XX
XX WPI; 2003-256589/25.
XX
XX
PT New expression vector containing gene(s) that encode a poly-gamma-
PT glutamate synthetase complex, useful for producing proteins (e.g.
PT vaccines or enzymes) on the microbial surface of Gram-positive and/or
PT Gram-negative bacteria.
XX
XX
PS Example 1; Page 113-120; 122pp; English.
XX
XX
CC The invention relates to a vector for expression of a target protein on a
CC microbial cell surface. The vector of the invention comprises either one
CC or more than two genes of the Bacillus subtilis poly-gamma-glutamate
CC synthetase complex (pgsBCA) to facilitate microbial surface expression of
CC the target protein. The pgsBCA gene complex comprises the pgsB, pgsB and
CC pgsA genes and is normally expressed in the outer membrane of Bacillus
CC subtilis. The vector can be transformed into either Gram-positive or Gram
CC -negative bacteria (e.g., Escherichia coli), and can be used for the
CC surface expression of various proteins of interest such as enzymes,
CC antigens, antibodies, attachment proteins or adsorption proteins.
CC Proteins recombinantly produced using the vector of the invention can be
CC used as, for example, vaccines or enzymes. The present sequence
CC represents a microbial cell surface expression vector of the invention,
CC pGNBCA.
XX
XX
SQ Sequence 6536 BP; 1855 A; 1464 C; 1567 G; 1649 T; 0 U; 1 Other;

Query Match 100.0%; Score 1182; DB 10; Length 6536;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTGTTACTCATTAATAGCTGCTGCTCATCTGATGCTGATCGGAATTTAGAAAAA 60
DB 224 ATGGGCTGTTACTCATTAATAGCTGCTGCTCATCTGATGCTGATCGGAATTTAGAAAAA 283
QY 61 CGACGACATCAGAAAAAATTGATGCCCTCCCTGTTCCGGGTGAATATTAAAGCGATCCGC 120
```

```

Db 284 CGACGACATCAGAAAAACATTCGATGCGCTCCCTGTCGGGTGATATTAAGCGATCGC 343
Qy 121 GGAATAATGACTGTGACAAAGGCTGACAAACCGGATATTTAATGAAAGCCGGTTACAAAGCT 180
Db 344 GGAATAATGACTGTGACAAAGGCTGACAAACCGGATATTTAATGAAAGCCGGTTACAAAGCT 403
Qy 181 GTTGAAAAACAACGAAACAGATGCAAGATATTTACTGGAGACACACCGGAGAAAAAG 240
Db 404 GTTGAAAAACAACGAAACAGATGCAAGATATTTACTGGAGACACACCGGAGAAAAAG 463
Qy 241 CCGATTTAAACGGAAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAGTCAATGAGAGA 300
Db 464 CCGATTTAAACGGAAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAGTCAATGAGAGA 523
Qy 301 ACAGTAGAAAGAGGGGCTTAACGCGATTTGTCAGTGAATGCAATGCGCTGTTAACCCAGATTAT 360
Db 524 ACAGTAGAAAGAGGGGCTTAACGCGATTTGTCAGTGAATGCAATGCGCTGTTAACCCAGATTAT 583
Qy 361 CAATATCATTTTTCAGAGAAACTTCTGCGAGGCAATATGCGGCTCATTTGTAATGTTTAA 420
Db 584 CAATATCATTTTTCAGAGAAACTTCTGCGAGGCAATATGCGGCTCATTTGTAATGTTTAA 643
Qy 421 GAAGACCATATGAGATGTCATGGGGCCGAGCGTTGATGAATTCAGAAAGCGTTTACCGCT 480
Db 644 GAAGACCATATGAGATGTCATGGGGCCGAGCGTTGATGAATTCAGAAAGCGTTTACCGCT 703
Qy 481 ACATTTCTTATATATGCGCATCTTGTCAATTAAGATGATGAATATACGAGTTCTTTTAA 540
Db 704 ACATTTCTTATATATGCGCATCTTGTCAATTAAGATGATGAATATACGAGTTCTTTTAA 763
Qy 541 CAAAAAGCAAAAGAACGAAACACAAAAGTCATATGCTGATTAACCTCAAAAATTTAAGAT 600
Db 764 CAAAAAGCAAAAGAACGAAACACAAAAGTCATATGCTGATTAACCTCAAAAATTTAAGAT 823
Qy 601 GAGTATTTACGTATTTTGAATACATGATATTCCTGATTAACGCTTCTGCGCGCTGGGT 660
Db 824 GAGTATTTACGTATTTTGAATACATGATATTCCTGATTAACGCTTCTGCGCGCTGGGT 883
Qy 661 GTGGCTCAAGCACTCGGCAATTCAGAAAGAACGATTTAAGGAAATGCTGAATGCGCG 720
Db 884 GTGGCTCAAGCACTCGGCAATTCAGAAAGAACGATTTAAGGAAATGCTGAATGCGCG 943
Qy 721 CCAGATCCGGAGCAATGAGAAATTTCTCGCTGATACGTCGAGCGAGCTGGGCACTTT 780
Db 944 CCAGATCCGGAGCAATGAGAAATTTCTCGCTGATACGTCGAGCGAGCTGGGCACTTT 1003
Qy 781 GTTATGCGTTTCCCGCAACGAGCGCTTCTTACTGTAATATGAAAGCGTAAAA 840
Db 1004 GTTATGCGTTTCCCGCAACGAGCGCTTCTTACTGTAATATGAAAGCGTAAAA 1063
Qy 841 GAAATCGTTTACCCGACCGATGATCCGATCATCATGAACTGCGCGAGACCGGTGC 900
Db 1064 GAAATCGTTTACCCGACCGATGATCCGATCATCATGAACTGCGCGAGACCGGTGC 1123
Qy 901 GATCGGACACAGCAATTCGCAAAATGACGTAATTCCTTATATTAAGCAAGTGAATC 960
Db 1124 GATCGGACACAGCAATTCGCAAAATGACGTAATTCCTTATATTAAGCAAGTGAATC 1183
Qy 961 TTTATCGGTGAACAACAGAACCGATGCTTAAAGCCATGAAAGGCAAAATTCCTGCA 1020
Db 1184 TTTATCGGTGAACAACAGAACCGATGCTTAAAGCCATGAAAGGCAAAATTCCTGCA 1243
Qy 1021 GACAAATGCGATGACCTAGAGATTAAGTCAACAGATGAATTAATGAAATTTGTTAAAGAAA 1080
Db 1244 GACAAATGCGATGACCTAGAGATTAAGTCAACAGATGAATTAATGAAATTTGTTAAAGAAA 1303
Qy 1081 AGAATGCAACCGGTGTCATATATGCGGTGCGCAATATTCATGCTGCGAGAGCTTTA 1140
Db 1304 AGAATGCAACCGGTGTCATATATGCGGTGCGCAATATTCATGCTGCGAGAGCTTTA 1363
Qy 1141 ATTGAAAAATCCAGGAATCAAGGTAAGAGAGTTCCTTAAGC 1182
Db 1364 ATTGAAAAATCCAGGAATCAAGGTAAGAGAGTTCCTTAAGC 1405

```

```

RESULT 5
AAE82255
ID AAE82255 standard; DNA; 1182 BP.
XX
AC AAE82255;
XX
DT 21-JUN-2001 (first entry)
XX
DE Bacillus subtilis ffo 336 DNA encoding a PGA synthesizing enzyme.
XX
KW Bacterial; poly-gamma-L-glutamic acid; L-PGA;
XX poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; db.
OS Bacillus subtilis.
FH
FT Key Location/Qualifiers
FT CDS 1..1182
FT /*tag= a
PN JP2001017182-A.
PD 23-JAN-2001.
XX
PF 09-JUL-1999; 99JP-00196335.
XX
PR 09-JUL-1999; 99JP-00196335.
XX
PA (NAGS ) NAGASE SANGYO KK.
XX
DR WPI: 2001-285408/30.
DR P-PSDB; AAE74024.
XX
PT New nucleic acid encoding a glutamate racemase enzyme useful for the
PS preparation of poly-gamma-glutamic acid.
XX
PS Claim 2, Page 11-12; 17pp; Japanese.
XX
CC The present sequence encodes an enzyme which is useful in the production
CC of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L-glutamic acid
CC (PGA). A plasmid comprising the present sequence may be used to transform
CC Escherichia coli. The transformants express the enzyme and PGA is
CC produced in the culture
XX
SQ Sequence 1182 BP; 399 A; 235 C; 272 G; 276 T; 0 U; 0 Other;
Query Match 99.1%; Score 1171.2; DB 4; Length 1182;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 7 TGGTACTCATATTAAGCTGTGCTGTCTGATCATGTCATGCGAATATTAAGAAAAACAGCA 66
Db 4 TGGTACTCATATTAAGCTGTGCTGTCTGATCATGTCATGCGAATATTAAGAAAAACAGCA 63
Qy 67 CATCAGAAAAACATTATATGCTTCTTCTGCGGTGAATTAAGCGCATTCGCGGAGAAA 126
Db 64 CATCAGAAAAACATTATATGCTTCTTCTGCGGTGAATTAAGCGCATTCGCGGAGAAA 123
Qy 127 TCGACTGTGCAAGGCTGACAAACCGGAATATTAATGAAGCCGTTACAAGACTGTTGA 186
Db 124 TCGACTGTGCAAGGCTGACAAACCGGAATATTAATGAAGCCGTTACAAGACTGTTGA 183
Qy 187 AAAACAACAGAAACAGATGCAAGATGATTTACTGGACACACCGGAGAAAAAGCGATT 246
Db 184 AAAACAACAGAAACAGATGCAAGATGATTTACTGGACACACCGGAGAAAAAGCGATT 243
Qy 247 AAAAGGAACCTTAGGGGCGGAATATCGAGAGCAAAAAGAGTCAATGAGAAACAGTA 306
Db 244 AAAAGGAACCTTAGGGGCGGAATATCGAGAGCAAAAAGAGTCAATGAGAAACAGTA 303
Qy 307 GAAAGAGGGGCTTAACCGATGTCAGTGAATGATGCTGTTAACCCAGATTATCAATC 366

```

```
Db 304 GAAAGAGGGCTAAACGCATTTGTCAAGTATGATGATGCTGTTAAACCAATTAATCAAAATC 363
Qy ATCTTTGAGAAAGAACTTCTGAGGCGCAATATCGGCGTCAATTGTGTAATGTTTGAAGAC 426
Db ATCTTTGAGAAAGAACTTCTGAGGCGCAATATCGGCGTCAATTGTGTAATGTTTGAAGAC 423
Qy CATATGATGTCATGAGGCGCAACGCTTGATGAATTTGAGAAAGCGTTTACCGCTACAAAT 486
Db CATATGATGTCATGAGGCGCAACGCTTGATGAATTTGAGAAAGCGTTTACCGCTACAAAT 483
Qy CCTTATTAATGCGCATCTTGTCTTAACAGATAGTGAATTAACCGAGTTCTTTAAACAAAA 546
Db CCTTATTAATGCGCATCTTGTCTTAACAGATAGTGAATTAACCGAGTTCTTTAAACAAAA 543
Qy GCAAAAGAACGAAACAAAAAGTCATCATCTGCTGATACTCAAAAATTACAGATAGTAT 606
Db GCAAAAGAACGAAACAAAAAGTCATCATCTGCTGATACTCAAAAATTACAGATAGTAT 603
Qy TTAAGTAATTTTGAATACATGATATTCCTGATAAACGCTTCTGCGCGCTGGGTGAGCT 666
Db TTAAGTAATTTTGAATACATGATATTCCTGATAAACGCTTCTGCGCGCTGGGTGAGCT 663
Qy CAAGCACTCGGCATGAGAAAGAAACAGCATTTAAAGGAAATGCTGAATGCGCCGCAGAT 723
Db CAAGCACTCGGCATGAGAAAGAAACAGCATTTAAAGGAAATGCTGAATGCGCCGCAGAT 727
Qy CCGGAGAGCAATGAGAAATCTTCCGCTGATCCGAGCGAGCCCTGGGACCTTTGTTAAT 786
Db CCGGAGAGCAATGAGAAATCTTCCGCTGATCCGAGCGAGCCCTGGGACCTTTGTTAAT 783
Qy GGGTTTGGCGGAAACGACGCTTCTTCTAATTGAAATATAGAAACGTTAAAGAAATC 846
Db GGGTTTGGCGGAAACGACGCTTCTTCTAATTGAAATATAGAAACGTTAAAGAAATC 843
Qy GGTATCCCGAGCGATGATCCGATCATCATGAACTCCGCGAGACCGCTGTGATGG 906
Db GGTATCCCGAGCGATGATCCGATCATCATGAACTCCGCGAGACCGCTGTGATGG 903
Qy ACACAGCAATTCGCAAAATGACGATATTCGCTTAATTTGAAGCAAGTGAATGATCTTAATC 966
Db ACACAGCAATTCGCAAAATGACGATATTCGCTTAATTTGAAGCAAGTGAATGATCTTAATC 963
Qy GGTGAAACAAAGAAACCGATGCTAAAGCCTATGAAAGAGCAAAATTCCTGACAGAA 1026
Db GGTGAAACAAAGAAACCGATGCTAAAGCCTATGAAAGAGCAAAATTCCTGACAGAA 1023
Qy CTGCTAGCTTAAGATTAATGATCAACAGATGAATTTATGAAATTTTAAAGAAATG 1086
Db CTGCTAGCTTAAGATTAATGATCAACAGATGAATTTATGAAATTTTAAAGAAATG 1083
Qy CACAAACGCTGATATGAGGCGTGGCAATATTCATGAGCGCGAGACCTTTAATGAA 1146
Db CACAAACGCTGATATGAGGCGTGGCAATATTCATGAGCGCGAGACCTTTAATGAA 1143
Qy AAAATCCACGAATACAGATTAAGCAAGCTCGTAAAC 1182
Db AAAATCCACGAATACAGATTAAGCAAGCTCGTAAAC 1179

RESULT 6
AAF82254
ID AAF82254 standard; DNA; 3045 BP.
XX
AC AAF82254;
XX
XX 21-JUN-2001 (first entry)
XX
Db Bacillus subtilis fpo 3336 DNA encoding glutamate racemase enzyme.
XX
XX Bacillus subtilis fpo 3336 DNA encoding glutamate racemase enzyme.
KM Bacterial: poly-gamma-L-glutamic acid; L-PGA;
XX poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; ds.
XX
.OS Bacillus subtilis.
```

```
XX Key Location/Qualifiers
FH 19..1200
FT CDS /*tag= a
FT CDS 1215..1264
FT CDS /*tag= b
FT CDS 1683..2825
FT CDS /*tag= c
PN JP2001017182-A.
XX
XX 23-JAN-2001.
PD
XX
XX 09-JUL-1999; 99JP-00196335.
PF
XX 09-JUL-1999; 99JP-00196335.
PR
XX (NAGS ) NAGASE SANGYO KK.
PA
XX WPI; 2001-285408/30.
DR P-PSDB; AAB74024, AAB74025, AAB74026.
DR
XX
XX New nucleic acid encoding a glutamate racemase enzyme useful for the
PT preparation of poly-gamma-glutamic acid.
XX
XX Claim 1; Page 7-11; 17pp; Japanese.
PS
XX
XX The present sequence encodes a glutamate racemase enzyme which is useful
CC in the production of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L
CC -glutamic acid (PGA). A plasmid comprising the present sequence may be
CC used to transform Escherichia coli. The transformants express the enzyme
CC and PGA is produced in the culture
XX
XX Sequence 3045 BP; 1000 A; 593 C; 696 G; 756 T; 0 U; 0 Other;
SQ
Query Match 99.1%; Score 1171.2; DB 4; Length 3045;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 7 TGGTACTCAATTAAGCCTGCTGCTCATATCTGTCATCGGAATATTAGAAAAAGACGA 66
Db TGGTACTCAATTAAGCCTGCTGCTCATATCTGTCATCGGAATATTAGAAAAAGACGA 81
Qy 67 CATCAGAAAAACATTTGATGCTTCCCTGCTGCTGATATTTAAACGCGATCCGCGAAAA 126
Db CATCAGAAAAACATTTGATGCTTCCCTGCTGCTGATATTTAAACGCGATCCGCGAAAA 141
Qy 127 TCGACTGTGACAGGCTGACCAACCGGAATATTATAGAGCCGCTTCAAGACTGTTGGA 186
Db TCGACTGTGACAGGCTGACCAACCGGAATATTATAGAGCCGCTTCAAGACTGTTGGA 201
Qy 187 AAAACAACAGGAACAGATGCAAGATATTACTGAGGACACACCGGAGGAAAAACCGATT 246
Db AAAACAACAGGAACAGATGCAAGATATTACTGAGGACACACCGGAGGAAAAACCGATT 261
Qy 202 AAAACAACAGGAACAGATGCAAGATATTACTGAGGACACACCGGAGGAAAAACCGATT 261
Qy 247 AAACGAAAACTCAGGGGCGCAATATCGAGAGCAAAAGAAAGTCAATGAGAAACAGTA 306
Db AAACGAAAACTCAGGGGCGCAATATCGAGAGCAAAAGAAAGTCAATGAGAAACAGTA 321
Qy 262 AAACGAAAACTCAGGGGCGCAATATCGAGAGCAAAAGAAAGTCAATGAGAAACAGTA 321
Qy 307 GAAAGAGGGGCTAACCGCATTTGATGATGATGATGCTGTTAAACCAATTAATCAAAATC 366
Db GAAAGAGGGGCTAACCGCATTTGATGATGATGATGCTGTTAAACCAATTAATCAAAATC 381
Qy 322 GAAAGAGGGGCTAACCGCATTTGATGATGATGATGCTGTTAAACCAATTAATCAAAATC 381
Qy 367 ATCTTTGAGAAAGAACTTCTGAGGCGCAATATCGGCGTCAATTGTGTAATGTTTGAAGAC 426
Db ATCTTTGAGAAAGAACTTCTGAGGCGCAATATCGGCGTCAATTGTGTAATGTTTGAAGAC 441
Qy CATATGATGTCATGAGGCGCAACGCTTGATGAATTTGAGAAAGCGTTTACCGCTACAAAT 486
Db CATATGATGTCATGAGGCGCAACGCTTGATGAATTTGAGAAAGCGTTTACCGCTACAAAT 501
Qy CCTTATTAATGCGCATCTTGTCTTAACAGATAGTGAATTAACCGAGTTCTTTAAACAAAA 546
Db CCTTATTAATGCGCATCTTGTCTTAACAGATAGTGAATTAACCGAGTTCTTTAAACAAAA 546
```



```

Db      502 CCTTAAATGCGCATCTTGTCTATTAACAGATAGTGAATATACGAGTCTTTAAACAAA 561
Qy      547 GCAGAAAGACGAAACACAAAAGTCATCATCTGATTAACCTCAAAATTAACAGATGATAT 606
Db      562 GCAGAAAGACGAAACACAAAAGTCATCATCTGATTAACCTCAAAATTAACAGATGATAT 621
Qy      607 TTACGTAATTTTGAATACATGATATCTCCCTGATTAACGCTTCTCCGCGCTGGGTGTGCT 666
Db      622 TTACGTAATTTTGAATACATGATATCTCCCTGATTAACGCTTCTCCGCGCTGGGTGTGCT 681
Qy      667 CAAGACCTCGGATGACGAGAGAAACAGATTTAAGGAATCTGAATCCGCGCAGAT 726
Db      682 CAAGACCTCGGATGACGAGAGAAACAGATTTAAGGAATCTGAATCCGCGCAGAT 741
Qy      727 CCGGAGACATGAGATTTCTTCCGCTGATCACTCCGAGGAGCCTGGGCACTTTTAT 786
Db      742 CCGGAGACATGAGATTTCTTCCGCTGATCACTCCGAGGAGCCTGGGCACTTTTAT 801
Qy      787 GGGTTTCCGCAACGACGCTTCTTCTACTTTGAATATATGAAACGCTTAAAGAAATC 846
Db      802 GGGTTTCCGCAACGACGCTTCTTCTACTTTGAATATATGAAACGCTTAAAGAAATC 861
Qy      847 GGTACCCGACCGATGATCCGATCATCATGAACTGCGCGACAGCCGTGCTGATCGG 906
Db      862 GGTACCCGACCGATGATCCGATCATCATGAACTGCGCGACAGCCGTGCTGATCGG 921
Qy      907 ACACAGCAATTCGCAATGACGTAATTCCTTATATTTGAAGCAAGTGAATGATCTTATC 966
Db      922 ACACAGCAATTCGCAATGACGTAATTCCTTATATTTGAAGCAAGTGAATGATCTTATC 981
Qy      967 GGTGAAACAAACGAAACCGATCTGTAAGAGCTTGAAGAGCAAAATTCCTGACAGAA 1026
Db      982 GGTGAAACAAACGAAACCGATCTGTAAGAGCTTGAAGAGCAAAATTCCTGACAGAA 1041
Qy      1027 CTGCTGACCTGAGATGATGATCAACAGATGAAATTTGTAAGAAAGAAATG 1086
Db      1042 CTGCTGACCTGAGATGATGATCAACAGATGAAATTTGTAAGAAAGAAATG 1101
Qy      1087 CACAACCGTGTATATATGCGCTGCGCAATTCATGATGCGCAGAGCCTTTAATTGAA 1146
Db      1102 CACAACCGTGTATATATGCGCTGCGCAATTCATGATGCGCAGAGCCTTTAATTGAA 1161
Qy      1147 AAATTCACGAAATACAGGTAAGAGCTGTAAGC 1182
Db      1162 AAATTCACGAAATACAGGTAAGAGCTGTAAGC 1197

RESULT 7
ABK75403
ID      ABK75403 standard; DNA; 819 BP.
XX      ABK75403;
XX      13-AUG-2002 (first entry)
XX      Bacillus licheniformis genomic sequence tag (GSTR) #2694.
XX      Differential gene expression; genomic sequenced tag; GSTR;
XX      altered culture condition; environmental stress;
XX      physiological provocation; ds.
XX      Bacillus licheniformis.
XX      WO200229113-A2.
XX      11-APR-2002.
XX      05-OCT-2001; 2001WO-US031437.
XX      06-OCT-2000; 2000US-00680598.
XX      27-MAR-2001; 2001US-0279526P.
XX      (NOVO ) NOVOZYMES BIOTECH INC.

```

```

PA      (NOVO ) NOVOZYMES AS.
XX      Berka R, Clausen IG;
XX      WPI; 2002-416684/44.
DR      Monitoring differential expression of several genes in first Bacillus
XX      PT cell relative to expression of same genes in one or more second Bacillus
XX      PT cells, by using substrate containing Bacillus genomic sequenced tag
XX      array.
PS      Claim 4; SEQ ID NO 2694; 200bp; English.
XX      The invention describes a method of monitoring differential expression of
XX      genes in a first Bacillus cell relative to expression of the genes in
XX      other Bacillus cells, comprising hybridising labeled nucleic acid probes
XX      isolated from Bacillus cells to a substrate containing array of Bacillus
XX      genomic sequenced tags (GSTR), examining the array, and determining
XX      relative gene expression by an observed hybridisation reporter signal of
XX      a spot in the array. The method is useful for measuring the expression of
XX      genes in a first Bacillus cell relative to expression of the same genes
XX      in one or more second Bacillus cells. The method is useful for monitoring
XX      global expression of several genes from a Bacillus cell, discovering new
XX      genes, identifying possible functions of unknown open reading frames and
XX      monitoring gene copy number variation and stability. Monitoring changes
XX      in expression of genes may be used to provide a representation of the way
XX      in which Bacillus cells adapt to changes in culture conditions,
XX      environmental stress or other physiological provocation. Extensive follow
XX      up characterisation is unnecessary, when one spot on an array equals one
XX      gene or one open reading frame, since sequence information is available.
XX      This sequence represents a genomic sequence tag (GSTR) used in the method
XX      of the invention. Note: The sequence data for this patent did not form
XX      part of the printed specification, but was obtained in electronic format
XX      directly from WPI at ftp.wpi.int/pub/published_pct_sequences

SQ      Sequence 819 BP; 257 A; 166 C; 215 G; 180 T; 0 U; 1 Other;
Query Match      46 4%; Score 548.4; DB 6; Length 819;
Best Local Similarity 79.9%; Pred. No. 2.3e-144;
Matches 645; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
Qy      24 CTGTGCTGTCATACGTCATCGATATTTAGAAAACGACGACATCAGAAAACATTTGA 83
Db      4 CTGTGTCATGTCGTGGGATCGCATTTATGAAAAAGGGCCACGACAAATATGCA 63
Qy      84 TGCCCTCCCTGTTCGGGTGAATATTTAAGCGCATTCGCGGAAATGACCTGTGACAAAGCT 143
Db      64 TGCGTGCCTGTCCGATGATCATCAACGATATACGCGGAAAGTCCACGCTGACAAAGAT 123
Qy      144 GACACCGGAATTTATTTAGAAACCGGCTTACAAAGCTGTGAAAAACAACGAGAAACAGA 203
Db      124 AACACAGGAGATTTATTTAGAAACAGGCTTACAAACAGTAGAANAACAACCGGACAGA 183
Qy      204 TGCAAGATGATTTACTGGGACACACCGGAGAAAAAGCGATTAAACGAAAACCTCAGGG 263
Db      184 CGAGAGATGATTTATTTAGGACACACCGGAGAAAGCGATCAAAAGAAAGCGGCAAG 243
Qy      264 GCGGATATCGAGAGCAAAAAGAACTCATGAGAGAAACAGTAGAAAGAGGGCTTAACGC 323
Db      244 GCGGATATTCGAGAGCAAGAGAGGTATGAAAGAAACGTGAGAAAGAGGGCCCAATGC 303
Qy      324 GATTGTCAGTGAATGATGCTGTTTAAACCAATTTCAATCATCTTTCCAGAGAAAGCT 383
Db      304 GATTGTCAGTGAATGATGCTGTTTAAACCAATTTCAATCATCTTTCCAGAGAAAT 363
Qy      384 TCTGACAGGCGCAATATGCGGCTCATTTGTAATGTTTAAAGAACATATGATGATGCG 443
Db      364 GCTTACAGGCTTAATATGCGGCTCATTTGTAATGTTTAAAGAACATATGATGATGCG 423
Qy      444 GCCGACCTTGATGAATTTGACAGAGCGTTTACCGGTACAAATTCCTTATATGCGCATCT 503
Db      424 ACCGACCTTGATGAATTTGACAGAGCGTTTACCGGTACAAATTCCTTATATGCGCATTT 483

```


QY	504	IGTCACTTACAGATPAGTAATATACCGAGCTCTTTAAACAAAAAGCAAAAGCAAC	563
Db	484	GGTATTACTGTAGTAGATATACCGATTTCTTTAAGCAAAATGGCAAAAGAAACAC	543
QY	564	AAAGATCATCATTCGTGTGATTAACCTCAAAAATTAACAGATGAGTATTTACGTAATTTTGAAT	623
Db	544	AAAGATCATTCGTGTGATTAACCTCAAAAATTAACAGATGAGTATTTACGTAATTTTGAAT	603
QY	624	CATGTTATTTCCCTGATTAACGCTTTCTGTGGCGCTGGTGTGGCTCAACGACTCGGCAATTGA	683
Db	604	CATGTTATTTCCCTGATTAACGCTTTCTGTGGCGCTGGTGTGGCTCAACGACTGGGCAATTGA	663
QY	684	CGAAGAAACAGATTTAAGGGGAATGCTGAATGCGCGCAAGATCCGGGAGCAATGGAAT	743
Db	664	CGAAGAAACAGATTTAAGGGGAATGCTGAATGCGCGCAAGATCCGGGAGCAATGGAAT	723
QY	744	TCTTCGCGTGAATCAGTCCGAGCGAGCTGTGGGCACTTTGTAAATGAGTGTGGCCGCAACGA	803
Db	724	TCTTCGCGTGAATCAGTCCGAGCGAGCTGTGGGCACTTTGTAAATGAGTGTGGCCGCAATGA	783
QY	804	CGCTTCTTCTACTTTGAAATATATGAA	830
Db	784	CGAGCTTCCACTTTAAACATTTGGAA	810
RESULT 8			
ABN91998	ABN91998 standard; DNA; 1164 BP.		
ID	ABN91998	standard; DNA; 1164 BP.	
AC	ABN91998;		
XX	24-JUL-2002	(first entry)	
DT	24-JUL-2002	(first entry)	
DE	Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1461.		
DE	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;		
KW	antibacterial; gene therapy; gene; ds.		
XX	Staphylococcus epidermidis.		
OS	US6380370-B1.		
PN	30-APR-2002.		
PD	13-AUG-1998;	98US-00134001.	
XX	14-AUG-1997;	97US-0055779P.	
PR	08-NOV-1997;	97US-0064964P.	
XX	(GENO-) GENOME THERAPEUTICS CORP.		
PA	Doucette-Stamm LA, Bush D;		
PI	WPI; 2002-381255/41.		
DR	P-PsDB; ABP39453.		
XX	Novel isolated nucleic acid encoding a Staphylococcus epidermis		
PT	polypeptide, useful for diagnosing and treating bacterial infections.		
PS	Disclosure; SEQ ID NO 1461; 267bp; English.		
XX	ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading		
CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences		
CC	given in ABP3124 to ABP37960. The S. epidermidis sequences have		
CC	antibacterial activity and can be used in gene therapy. The sequences can		
CC	also be used in the diagnosis and treatment of bacterial infections,		
CC	particularly S. epidermidis infections. The sequences can be used to		
CC	screen for compounds able to interfere with the S. epidermidis life cycle		
CC	or inhibit S. epidermidis infection. N.B. The sequence data for this		
CC	parent did not form part of the printed specification, but was obtained		
CC	in electronic format directly from the USPTO web site		
SEQ	Sequence 1164 BP; 415 A; 172 C; 244 G; 333 T; 0 U; 0 Other;		

Query Match	28.6%	Score 338.6	DB 6	Length 1164
Beet Local Similarity	59.7%	Pred. No. 6e-85		
Matches 588	Conservative 0	Mismatches 394	Indels 3	Gaps 1
QY	5	GCTGGTACTACTATTATAGCCCTGTGCTG--TCATACTGTGTCATCGGAATATTATGAAAAAC	61	
Db	5	GATTGTTACTCATTTATATAGCGTGTGTGGCTTATTCTCTGGCTAGGAATTAAGAGAAGA	64	
QY	62	GACGACATCAGAAAAACATTGATGCCCTCCCTGTTCCGGGTGAATTTAAACGCATCCGCG	121	
Db	65	AGCCTCAGCAAAATCGACTTAAAAAAATTCATTACGTATTAACCTAAATGGAATTCGGT	124	
QY	122	GAATAATCGACTGTGACAAAGCTGCAACACCGAATATTATATGAAGCCGGTTTACAAGACTG	181	
Db	125	GGAATCTACGATTACTCGATGGCTTACAGTGTTCATACGTGAGGACCAAGTATACAGTTG	184	
QY	182	TTGGAAAAACAACAGGAAACAGATCGAAGATGATTTACTGGGACACACCGGAGAAAAAC	241	
Db	185	TTGGTAAAGACAACAGGAAACAGATCAGCAAAAGCTCTATTTGGTTTACTGAAAAAGAAATATC	244	
QY	242	CGATTAAACGGAAACCTCAGGGGCGCAATATCGAGACCAAAAAAGAACTCATGAGAGAA	301	
Db	245	CAGTATATGAAAAACACAGAGCGCCACATTTGGAGAACACAGATTTATTCTGAAG	304	
QY	302	CAGTAAAGAGGGGCTTAAACGCAATGTTCAGTGAATGCAATGCAATGCTTTAAACCAATATC	361	
Db	305	TGCTTAAACAAAAAGCAAAATGCCCTAGTTATAGGTATGATGGCTGTAAATCCAGATTTATC	364	
QY	362	AAATCATCTTTCAGAGAAACCTTCTGACGGCCAAATATCCGGGCTCATTCGTAATGTTTTAG	421	
Db	365	AAATTACCTTTCAAAAATGATTTAGTAAAGCAAAATATGGGTATATGTATTAATGTGATGG	424	
QY	422	AAGACCATATGATGTCAATGGGCGCGACGCTGTGAATGAATTCAGAAACGTTTACCGCTA	481	
Db	425	AAGACCATATGATGTCTTATAGACCGAACATTTAAAGATGTAGCGCAACTTTTACTGCA	484	
QY	482	CAATTCCTTATATATGCGCCATCTTGTCAATTACAGATAGTGAATATACGAGTTCTTTAAAC	541	
Db	485	CAATTCCTTATATATGCGGAAATTAGTTGTAAAGAAATACATATCTAGTTCTTTGTAA	544	
QY	542	AAAAAGCAAAAGAACGAAACAACAAAGCATCATTTGCTGATPAACCAAAAAATTACAGATG	601	
Db	545	AGGAAGCTTAAAAAGCGTATTCAGAACCTCATTTGTTAGTATTAAGACGTATACAGAAAT	604	
QY	602	AGTATTTACGTAAATTTTGAATACATGTGTATTCCTGATPAACGCTTCTCTGCGCTGGGTG	661	
Db	605	CATATTTTACGAAAGTTGCATTTATTTAGTATTTCCGATTAATGTAGCTATTGTTAGGAA	664	
QY	662	TGGCTCAAGCACTCGGCAATTGACGAAGAAACAGCAATTAAAGGAAATGCTGAATGGCGCG	721	
Db	665	TAGCGCAAGCAAGTTGGTGTATGAAGAAGAACTGCATTCAMAGTATGTTAAATGCAACAG	724	
QY	722	CAGATCCGGGAGCAATGAAATTCCTCCGCTGATCAAGTCGAGCGAGCGCTGGGCACTTGG	781	
Db	725	CCGATCCAGGTCGTGTTAGTAAATTAATTTCCATGCAAAATCGCACAAAAAATGATTTTG	784	
QY	782	TTAATGGGTTTGGCCGAAACGACGCTTCTTCTACTTTGAATATATATGAAACGTGTAAAG	841	
Db	785	TTAATGTCAATTCGTGCTTAAATACCGACGTCATCAAAAAGCAATTTTAAATAAAGTGGAAT	844	
QY	842	AAATCGGTTAACCCGACCGATGATCGATCATCATCATGAACTGCGCGCGACGCTGTGCG	901	
Db	845	CATATATTTATCCATACGATTAAGAAAAATATCATTTCTCATTTGCTGTCAAGTATAGGTTG	904	
QY	902	ATCGGACACAGCAATTCGCAAAATGACGTATTCCTTATATTGAAGCAAGTGAATGATCT	961	
Db	905	ATGAGAACCAACTCTTGTGTATTAATTTTATGGAAGTGAATGATTCATGATGTTCTCATTT	964	
QY	962	TAAATCGTGAACAAACAGAACCGAAT	986	
Db	965	GTACAGGAAAAAGTACCAAAATGGT	989	

RESULT 9
 ADS01519
 ID ADS01519 standard; DNA; 1164 BP.
 XX
 AC ADS01519;
 DT 04-NOV-2004 (first entry)
 XX
 DE Staphylococcus epidermidis polynucleotide seqid 814.
 XX
 KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
 KW recombinant expression vector; infection; computer readable medium;
 KW computer based system; gene; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN US2004147734-A1.
 PD 29-JUL-2004.
 PF 01-DEC-2003; 2003US-00724972.
 PR 08-NOV-1997; 97US-0064966A.
 PR 13-AUG-1998; 98US-00134001.
 PR 29-NOV-1999; 99US-00450969.
 XX
 PA (DOUC/) DOUCETTE-STAMM L.
 PA (BUSH/) BUSH D.
 PI Doucette-Stamm L, Bush D;
 DR WPI; 2004-580138/56.
 DR P-Psdb; ADS05291.
 PT New isolated polypeptide, and encoding nucleic acid derived from
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
 PT treating an S. epidermidis bacterial infection.
 XX
 PS Claim 5; SEQ ID NO 814; 741bp; English.
 CC The invention describes an isolated nucleic acid comprising a nucleotide
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
 CC given in the specification. Also described are: a recombinant expression
 CC vector; a cell comprising a recombinant expression vector of (1);
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection, comprising a nucleic acid cited above and a carrier; treating
 CC a subject for S. epidermidis infection; a recombinant or substantially
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a
 CC sample; a computer readable medium having recorded in it the nucleotide
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
 CC system for identifying fragments of the Staphylococcus genome of
 CC commercial importance; a computer based system for identifying fragments
 CC of the Staphylococcus plasmids of commercial importance; identifying
 CC commercially important nucleic acid fragments of the Staphylococcus
 CC genome and/or plasmids; and identifying an expression modulating fragment
 CC of the Staphylococcus genome and/or plasmids. The methods and
 CC compositions of the present invention are useful for the diagnosis,
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
 CC infection. This sequence encodes a S. epidermis protein of the invention.
 XX
 SQ Sequence 1164 BP; 415 A; 172 C; 244 G; 333 T; 0 U; 0 Other;

Query Match 28.6%; Score 338.6; DB 13; Length 1164;
 Best Local Similarity 59.7%; Pred. No. 66-85;
 Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;

OY 5 GCTGGTACTCAATTATAGCTGTGCTG---TCATACTGTCATCGGAATATAGAAAAAC 61
 Db 5 GATTTGTTACTCAATTATAGCTGTGCTGCTTATTTCTGAGCTAGAAATTAAGAGAGA 64
 OY 62 GAGGACATCAGAAAAACATTGATGCTTCCCTGTGGGGGGAATATTAAGCGATCCGGC 121
 Db 65 AGCGTCATGCAAAATCGACTGAAAAAAATTCATTACGATTAACATTAAGGAATTTGTC 124
 OY 122 GAAATGACCTGAGCAAGGCTGCAACCGGAATATTAATAGAACCGGTTTACAAGACTG 181
 Db 125 GGAATATCAGATTACTCGAGTGGCTTACAGTGTCTTACGTGAGGACCAAGATAGAGTTG 184
 OY 182 TTGGAATAAACAACAGAAACGATGCAAGATGATTTATCTGGGACACAACGAGGAAAAAC 241
 Db 185 TTGGTTAAGACAACAGAAACGATGCAAGATGCTTATTTGGTTTACTGAAAAAGAAATATC 244
 OY 242 CGATTAAACGAAACCTCAGGGGCGGAATATCGAGACAAAAAGAGTCTAGAGAGAA 301
 Db 245 CAGTAAATCAGAAACCAACAGGCGCCCAACATTTGAGAGAACACGAGATATTTGTAAGG 304
 OY 302 CAGTAAAGAGGGGCTTACCGGCAATTTGCAAGTGAATGATGGCTTTAACCAATTATC 361
 Db 305 TGGTTAAACAAAAAGCAAAATGCCCTAGTTAATGAGTGTATGCTGTAAATCAAGATTATC 364
 OY 362 AATCATCTTTCAGAAAGAACTTTCAGAGCCCAATATCGGCTGATTTGTGAATGTTTAA 421
 Db 365 AATTTACTTTCAAAATGATTTAGTAAAGCAATATCGGTATTTGTTAATGTGATGG 424
 OY 422 AAGACCATATGATGTCATGAGGGCGGACGCTTGATGAATGCAAGAGCTTTACCGCTA 481
 Db 425 AAGACCATATGATGTCATGAGGGCGGACGCTTGATGAATGCAAGAGCTTTTACTGCA 484
 OY 482 CAATTCCTTATTAAGGCGCATCTTGTCAATTAAGATGATATATACGAGTCTTTTAAAC 541
 Db 485 CAATTCCTTATTAAGGCGCAATTTAGTTGAATGAAGATATATATCTATCTTTCTTTGTA 544
 OY 542 AAAAAAGAAAGAAAGCAAAACAAAGATCATCTGCTGATTAACCAAAAAATTACAGATG 601
 Db 545 AGGAAGCTTAAAGAGGTAAATTCAGAACTCATTTGTTGATTAAGACGCTACACAGAT 604
 OY 602 AGTATTTACGTAATTTTGAATATATCATGTATTCCTGATTAACGCTTCTGCGGCTGGTG 661
 Db 605 CATATTTAAGGAAGTTGATTAATTTAGTATTTCCGATTAATGTAGCTATTTGTTAGGAA 664
 OY 662 TGGCTCAAGCACTCGGCAATTCAGAAAGAAACAGCATTTTAAGGAATGCTGAATGGCCGC 721
 Db 665 TAGCCCAAGCAAGTTGTGTAGTGAAGAACTGCAATTAAGATGTTTAATGACACAG 724
 OY 722 CAGATCCGGAGCAATGAAATTTCTTCGCTGATCAGTCCGAGCGAGCCTGGGCACTTTG 781
 Db 725 CCGATTCAGGTGCTGTGAATTTAAATTTCCATGCAAAATGCAAAAAATGATTTG 784
 OY 782 TTAATGGGTTTCCGCAACAGACGCTTCTTCACTTTGAATATATGAAAACTGTAAAG 841
 Db 785 TTAATGCAATGCTGCTGTATGAAACGCAAGTCAAAAGCAATTTTAATTAAGTGAAT 844
 OY 842 AATGCGTTACCCGACCGATGATCCGATCATATATGAATCTCCGCGCAACCGTGTG 901
 Db 845 CATATTAATATCATCATACGTAAAGAAATATATATCTCATTTGTGATAGGGTTG 904
 OY 902 ATCCGACACAGCAATTCGAAATGACGATTTGCTTATATTAAGCAAGTGAATGATCT 961
 Db 905 ATTAGAACACAACTCTTTGTTGAATCTTTTAGTGAAGTGAATTAAGATTTTCATTT 964
 OY 962 TATCGGTGAACACAGAACCGAT 986
 Db 965 GTACAGAAAAAGTACACAAATGT 989

RESULT 10
 AAH54379
 ID AAH54379 standard; DNA; 3240 BP.

AAH54379;
 03-SEP-2001 (first entry)
 S. epidermidis genomic polynucleotide sequence SEQ ID NO:3743.
 Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
 endocarditis; ds.
 Staphylococcus epidermidis.
 MO200134809-A2.
 17-MAY-2001.
 09-NOV-2000; 2000MO-US030782.
 09-NOV-1999; 99US-0164258P.
 (GLAXO) GLAXO GROUP LTD.
 Kimmerly WJ;
 MPI; 2001-316495/33.
 Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 useful for vaccinating against infections, e.g. endocarditis.
 Claim 8; Page 1366-1367; 2188pp; English.
 AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 (II), given in AM81454 to AM81454, from Staphylococcus epidermidis. (I)
 and (II) can have antibacterial activity and therefore can be used in
 vaccination. The nucleic acids (I) may be used to produce the S.
 epidermidis polypeptides (II) via the production of vectors containing
 them which are used to produce hosts cells which express the
 polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 used to vaccinate subjects and to raise antibodies against the bacteria.
 The polypeptides may also be used to assay for other inhibitors of their
 activity and therefore identify compounds that may be used for the
 treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 AAH55090 represent specifically claimed S. epidermidis genomic DNA
 polynucleotide sequences from the present invention. AAH55091 to AAH55098
 represent oligonucleotide sequences and primers which are used in the
 exemplification of the present invention. N.B. The present invention
 specifically claims all the polynucleotide sequences given in the
 sequence listing of the present specification, however the sequence
 listing only goes up to SEQ ID NO:4454 so even though sequences are given
 in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
 for SEQ ID NO:4455 to 4464
 Sequence 3240 BP; 1128 A; 464 C; 589 G; 1059 T; 0 U; 0 Other;
 Query Match 28.6%; Score 338.6; DB 4; Length 3240;
 Best Local Similarity 59.7%; Pred. NO. 9.5e-85;
 Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;
 5 GCTGTTACTATTAGCTGTGCTG--TCATCTGCTCATTCGGAATATTGAAAAAC 61
 1673 GATTGTACTACTATTAGCTGTGCTGCTGCTATTCTTTGGCTAAGAAATTAAGAGA 1732
 62 GACGACATCAGAAAAACATTGATGCCCTCCGCTGGGTGAATTTAAAGCAATCCGCG 121
 1733 AGCGTCATGCAAAATGACCTGAAAAAATTCATTACGATTAACATTAATGAAATTCGTG 1792
 122 GAAATCGACTGTGCAAGGCTGCAACCGGAATTTAATAGAACCGGTTTACAGACTG 181
 1793 GGAATCTACGATTCTCGGATGCTTACAGTCTTACGAGAGCAAGATTAAGTTG 1852
 182 TTGGAATAACAACAGAAACAGATGCAAGATGATTTACTGGGACACACCGAGGAAAAAC 241
 1853 TTGGTAAGACAACAGAAACAGATGCAAGATGCTCTATTGTTTACTGAAAAAGAAATATC 1912

242 CGATTAAACGGAACCTCAGAGGCGCAATATCGAGAGCAAAAAAGATCATAGAGAAA 301
 1913 CAGTAATCAGAAAAACCAAGAGGCGCAACATTTGAGAGAACAGAGATATTATTCGTAAG 1972
 302 CAGTAAAGAGGGGCTTAAGCGCATTTGTCAATGATGATGCTGTTAACCAGATTATC 361
 1973 TGGTAAACAAAAAGCAAAATGCCCTAGTTAATGATGATGCTGTTAATCCAGATTATC 2032
 362 AAATCATCTTTCAAGAAACCTTCGAGGCGCAATATCGGCGTCAATGATGATGATTTAG 421
 2033 AAATTCCTTTCAAAATGATTTAGTAAAGCAATATCGGCTGATTTAATGATGATG 2092
 422 AAGACATATGATGATGATGAGGCGGCGGCGCTGATGAAATGAGAAAGCGTTTACGCTA 481
 2093 AAGACATATGATGATGATGATGAGGCGGCGGCGCTGATGAAATGAGAAAGCGTTTACGCTA 2152
 482 CAATTCCTTTAATGCGCATCTTGTCTATTCAGATGATGATGATGATGATGATGATGATG 541
 2153 CAATTCATATTAACGGAATTAATGATGATGATGATGATGATGATGATGATGATGATG 2212
 542 AAAAGCAAAAGCAAAACCAAAAGTCATCATCTGCTGATGATGATGATGATGATGATG 601
 2213 AGAAGCTAAAGAGCTAATTCAGACTCATGTTGATGATGATGATGATGATGATGATGATG 2272
 602 AGTATTACGTAATTTGAAATACATGATGATGATGATGATGATGATGATGATGATGATG 661
 2273 CATATTACGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2332
 662 TGCTCAAGCACTGCGCATGAGCAAGAAACAGCATTTAAGGAATGATGATGATGATGATG 721
 2333 TAGGCAAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2392
 722 CAGATCCGGAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 781
 2393 CGATCCAGGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2452
 782 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 841
 2453 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2512
 842 AAATCGGTTACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 901
 2513 CATTAATTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2572
 902 ATGGAACAGCAATGCGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 961
 2573 ATGAACACCACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2632
 962 TAATCGGTAACCAAGAACCGAT 986
 2633 GTACAGAAAAAGTACCAATGAT 2657
 Db
 RESULT 11
 ID ADB06113
 ID ADB06113 standard; DNA; 1212 BP.
 AC ADB06113;
 20-NOV-2003 (first entry)
 Allolococcus oclitis antigenic protein encoding DNA SEQ ID NO:53.
 Allolococcus oclitis; antigenic protein; immunogenic; immunisation;
 gene therapy; Gram-positive Bacterium; infection; gene; ds.
 Allolococcus oclitis.
 MO2003048304-A2.
 12-JUN-2003.
 25-NOV-2002; 2002MO-US036123.

WP ADB12064_11 1100001 1210000
WP ADB12064_12 1200001 1310000
WP ADB12064_13 1300001 1410000
WP ADB12064_14 1400001 1510000
WP ADB12064_15 1500001 1610000
WP ADB12064_16 1600001 1710000
WP ADB12064_17 1700001 1754382
ID ADB12064 standard; DNA; 1754382 BP.
XX
XX ADB12064;
XX
XX 20-NOV-2003 (first entry)
XX
XX Allolococcus oclidis entire genome sequence SEQ ID NO:6651.
XX
XX Allolococcus oclidis; antigenic protein; immunogenic; immunisation;
XX gene therapy; Gram-positive bacterium; infection; gene; ds.
XX
XX Allolococcus oclidis.
XX
XX WO2003048304-A2.
XX
XX 12-JUN-2003.
XX
XX 25-NOV-2002; 2002WO-US036123.
XX
XX 29-NOV-2001; 2001US-0333777P.
XX 18-NOV-2002; 2002US-0426742P.
XX
XX (AMHP) WYETH HOLDINGS CORP.
XX
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX
XX WPI; 2003-505284/47.
XX
XX New Allolococcus oclidis polynucleotides and polypeptides, useful for
XX treating and diagnosing diseases, drug screening assays and monitoring of
XX effects during drug clinical trials.
XX
XX Example 3; SEQ ID NO 6651; 1019bp; English.
XX
XX The present invention describes an isolated polynucleotide (1) of
XX Allolococcus oclidis genomic DNA, which encodes an antigenic protein.
XX Allolococcus oclidis is a Gram-positive bacterium. Also described: (1)
XX an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
XX expression vector comprising the novel isolated polynucleotide (1), its
XX complement, degenerate variant or fragment; (3) a genetically engineered
XX host cell, transfected, transformed or infected with the vector of (2);
XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
XX composition comprising the polypeptide, its complement, biological
XX equivalent or fragment, or the polynucleotide that is comprised in the
XX expression vector; (6) a pharmaceutical composition comprising the
XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array
XX of the polypeptides of (1), their biological equivalent or fragment; (8)
XX immunising against Allolococcus oclidis by administering to a host the
XX immunogenic composition; (9) detecting and/or identifying Allolococcus
XX oclidis in the biological sample; (10) a kit comprising a container
XX containing the novel polynucleotide, its degenerate variant or fragment,
XX or the antibody of (4); and (11) producing a polypeptide by culturing the
XX genetically engineered host cell under conditions suitable to produce the
XX polypeptide from the culture. (1) can be used in gene therapy. The
XX polynucleotides, polypeptides, antibodies and compositions of the present
XX invention can be used for treating and diagnosing diseases, drug
XX screening assays and monitoring of effects during drug clinical trials.
XX The polynucleotides are useful for expressing and detecting Allolococcus
XX oclidis. The present sequence represents the entire genome of
XX Allolococcus oclidis, which is given in the exemplification of the
XX present invention.
XX
XX Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 00cher;
XX
XX Query Match 26.5%; Score 313.2; DB 9; Length 110000;
XX Best Local Similarity 55.1%; Pred. NO. 7,1e-77;

Matches 634; Conservative 0; Mismatches 513; Indels 3; Gaps 1;
QY 27 TCGTGTATATCGGTGATCGGAATATTGAAAAACGACATCAGAAAAATTGATGC 86
|||
DB 23405 TGGCTTCTTATTGGCTTACGCTTATTATGAAAGCAGCGTAAAGGAAATATTAAAA 23346
|||
QY 87 CTTCCCTGTTCCGGTGAATATTAAACGGCATCCGCGAAAAATGACCTGTGACAAAGCTGAC 146
|||
DB 23345 AATGACCTACGCAATTATTAATGACACAGAGGCAACAACTGACTCGTTGGT 23286
|||
QY 147 AACCGAATATTAAATAGAACCCGCTTCAAGACTCTTGGAAAAACACAGAACAGATGC 206
|||
DB 23285 TACTGGGGTCTACACACACCTGCTACCAAGTTATTGAAAAACACAGAACTGAGTC 23226
|||
QY 207 AAGAAATGATTACTGGGACACACCGAGAAAAAGCCGATTAAACGAAACCTCAGGGGCC 266
|||
DB 23225 ACGCATTAATTACTGGACACGAGCAGAAAGAGAGANTGAAAAGCCTCATCTGACCC 23166
|||
QY 267 GAATATCGAGAGCAAAAGAAAGTCAATGAGAGAAACGTAGAAAGAGGGCTTAACGCAT 326
|||
DB 23165 TAATATTCTGAACAATATCAAGTTATTGAAAAGGCAAGAAAAAGAGGGCTGATGCCCT 23106
|||
QY 327 TGTCAATGAAATGCATGGCTGTAAACCAATATCAATCATCTTACAGAAAGACTTCT 386
|||
DB 23105 TGTATTGAAATGTAATGCGAGTTATCCAGAGTACCAAGATGTTTATCTAATGAAATGTT 23046
|||
QY 387 GCAGGCAATATGCGGCGTCAATGTAATGTTTAAAGAACCATATGATGATGAGGCGC 446
|||
DB 23045 TCAATCAAAATTTACAGCCATCGTTAAGCTTTTGAAGACCAATGATATCTAGGTC 22986
|||
QY 447 GACGCTTGATGAATTCGAAAGCGTTTACCGCTACATCTTAAATGAGCCATCTGT 506
|||
DB 22985 AACAATGATCAATTCCTCGGGCTATGCAAGACCATCCCAAAATGAGCTGTGT 22926
|||
QY 507 CATTACAGATAGTAATATACCGAGTTCTTTAAACAAAACAAAAGAAACACAAA 566
|||
DB 22925 CATCTCAGATGATGACATTAAGACTTTTATTTGAAGTGCAGAGAAATTCGCG 22866
|||
QY 567 AGTATCATTTGCTGATTAACCTCAAAAATTACAGATGATATTACGTAATTTGAATCAT 626
|||
DB 22865 GGTAAAGTTATACGAAACCGACCGCTTAGTCAAGAAATTATGGATCGCTTTGACTATCT 22806
|||
QY 627 GGTATTCCTGATTAACCTTCTCTGGCGTGGTGTGCTCAAGACTCGGCATTTGACGA 686
|||
DB 22805 TGTCTTCCAAACAATAGCTGTTCGCTGGCTATTTGCGAAGAACTCGACGTTGACCG 22746
|||
QY 687 AGAAACAGCATTTAAGGAAATGCTGAATGCGCCGCGCATTCGGAGCAATGAAATTC 746
|||
DB 22745 CGATATCTGCCATTTAAAGTATGCTTCAAGCTGAACCTGATCCGGGCAATTATGCAATTGTT 22686
|||
QY 747 TCCGCTGATCAGTCCGAGCGAGCTGGGCACTT---GTTAATGGTTGCGCAACGA 803
|||
DB 22685 TGAATAGGGGAGGCTTATGACCAAGAGTTATTTAGCTTATGCTTGTGCTTAATGA 22626
|||
QY 804 CGCTTCTTCTAATTGAAATATATGAAAACGTAAAGAAATCGTTAACCCGACGATGA 863
|||
DB 22625 ACCGACTTTACATGCTGATCTATGAAAGCTCAAGAAAGTGGCTTAATAGGTGAAG 22566
|||
QY 864 TCCGATCATCATGAACTGCGCGCAACCGTGTGATCGACACAGCAATTCGAAA 923
|||
DB 22565 CCCATTATTATTGATCAATTGCCCCAGAAACCGGAGTGAACCGGCTAAATATTATTCACCA 22506
|||
QY 924 TGAGTATGCTCTTATATTGAAAGCAATGATCTTATGCGTGAACCAACAGAAC 983
|||
DB 22505 AGACTTATATGCTTAATTAATAAGATGCGAGGTTATGCCATTGGAACAAATATCGCAT 22446
|||
QY 984 GATCGTAAAAAGCCATGAAAGGCAAAATTCCTGACAGACAACTGCATGACCTAGAGTA 1043
|||
DB 22445 TATTGAGAGGCTATTAATATCAGTATAGATCAAGGATTAATTATCTTTGGAAG 22386
|||
QY 1044 TAACTCAACAGATGAATATGAAATGTTAAAGAAAAAAGATGACAAACCGTGCATATA 1103
|||
DB 22385 GAAATCTTGGAAAGATATTATACATGACCCCTAATGACTTATGCTGTGAAAGTTGCTT 22326
|||

QY 1104 TGGCGTCGCAATATTCATGTCGCCGACAGCCTTTAATGAAAAATCCAGAAATACA 1163
DB 22325 GGGATTTGGAATATTCATGTCGTGAGAGATTTTTCCTTACCTTGATGAAAAATTA 22266
QY 1164 GGTAAAGCAG 1173
DB 22265 CTTGTGGCAG 22256

RESULT 13

ADR48536
ID ADR48536 standard; DNA; 560 BP.

AC ADR48536;
DT 04-NOV-2004 (first entry)
XX capB target sequence #3.
DE
XX Bacillus anthracis; cutaneous; respiratory anthrax infection; pagA; capB;
KW pXO1; pXO2; target sequence; ds.
XX
OS Bacillus anthracis.
PN MO2004070001-A2.
XX
PD 19-AUG-2004.
XX
PF 12-NOV-2003; 2003WO-US036240.
XX
PR 15-NOV-2002; 2002US-0426552P.
PR 16-MAY-2003; 2003US-0471082P.
XX
PA (GENP-) GEN-PROBE INC.

PI Norman SA, Bungo JU, Hogan JU, Weisburg WG;
DR WPI; 2004-604428/58.
XX
XX

PT New oligonucleotides that hybridize specifically to a Bacillus anthracis
PT sequence, useful for detecting cutaneous and respiratory Bacillus anthrax
PT infections.

Claim 1; SEQ ID NO 34; 61bp; English.

CC The present invention relates to an oligonucleotide of 20-40 nucleotides
CC that specifically hybridizes to a sequence contained in a Bacillus
CC anthracis target sequence. The methods and compositions of the present
CC invention are useful for detecting the presence of Bacillus anthracis
CC nucleic acid in a sample, in particular for detecting cutaneous and
CC respiratory anthrax infections. Two synthetic genetic target sequences,
CC derived from pagA and capB gene sequences, were synthesized to provide
CC known standards for testing oligonucleotides for detection of the genes
CC carried by the plasmids pXO1 and pXO2, without requiring handling of
CC virulent Bacillus anthracis. The present sequence represents a capB
CC target sequence.
XX
XX

Sequence 560 BP; 181 A; 89 C; 131 G; 159 T; 0 U; 0 Other;

Query Match 24.9%; Score 294.2; DB 13; Length 560;
Best Local Similarity 70.6%; Pred. No. 1.6e-72;
Matches 392; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 190 ACAACGAGAACATGCAAGATGATTTACTGGAGACACCCGAGAAAAAGCGATTAA 249
DB 1 ACAACGAGAACATGCAAGATGATTTACTGGAGACACCCGATTAA 60
QY 250 CGGAACCTCAGGGGCCGAATATCGAGAGCAAAAAGTCAATGAGAAACAGTAGAA 309
DB 61 CCGCGTAAAGAGAGTCTTAATATCGGTGACCAACGAGGTACTTAAAGAGCTGCTGAT 120
QY 310 AGAGGGGCTTAACGCGATTGTCTAGTGAATGACGCTGTTAACCAAGATTATCAATATC 369

DB 121 TTAGAGAGAGACACTTATTTGTAATGATGACAGTTCAACCCGATTATCAATATATC 180
QY 370 TTTCAGAGAGACCTTCGACAGCCCAATATCGCGCTCATTTGTAATGTTTGAAGACAT 429
DB 181 TTCCAAATATAAATGATTCAGCAAAATGTTGAGTGAATGTTTGAAGATCAT 240
QY 430 ATGATGTCATGGGGCCGACCTTGATGAAATTCAGAGCGTTTACCGCTACAAATTCCT 489
DB 241 ATGATGTTATGAGACCTTACCTTGACAGAGTACTGAAAGCTTTCACGTCACTTCCA 300
QY 490 TATATGCGCATCTGTCATTACAGATGATGATATACCGAGTTCTTAAACAAAGCA 549
DB 301 TATATGACATTTAGTACATGATGAAAGTGAATTCCTTGATTAAGAGATTGCA 360
QY 550 AAAGAGCAAAACAAAAAGTCAATCATCTGCTGATTAACCAAAATTAAGATGATTTA 609
DB 361 GAAGAGCAAAATACAAAGATGATTTGCGGATATTTCAAAATTCAGAAATTTCTTA 420
QY 610 CGTATTTTGAATCATGTAATTCCTGATTAACGCTTCTGCGCTGCGTGGCTCA 669
DB 421 CGAAATTTGATTTACATGCTCTTCCAGATATATGATGCTTGTTCATTAGCGTAGCAGAG 480
QY 670 GCACTCGGCATTGACGAGAAACAGCATTTAAGGAAATGCTGAATGCGCCGAGATCG 729
DB 481 GCTCTGGGATTGATGAGAAACAGCATTCGCTGATTTGAATGCTCATCCGATCA 540
QY 730 GGAGCAATGAGAAAT 744
DB 541 GGAGCAATGAGAAAT 555

RESULT 14

AAH53464
ID AAH53464 standard; DNA; 861 BP.

AAH53464;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2321.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW endocarditis; ds.

OS Staphylococcus epidermidis.

PN MO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US030782.

PR 09-NOV-1999; 99US-0164258P.

PA (GLAXO) GLAXO GROUP LTD.

PI Kimmery WJ;

DR WPI; 2001-316495/33.

DR P-PSDB; AAG82614.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.

PS Claim 8; Page 616-617; 2188bp; English.

CC AAH53304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAH81454 to AAH83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX

SQ Sequence 861 BP; 311 A; 123 C; 175 G; 252 T; 0 U; 0 Other;

Query Match 20.8%; Score 246.2; DB 4; Length 861;

Best Local Similarity 60.0%; Pred. No. 7.3e-59;

Matches 410; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

```
QY 304 GTAGAAAAGGGGCTAAGCGATGTCAGTGAATGCAATGCGCTGTTAAACCGATATCA 363
    |||
DB 4 GTTAAACAAAAGCAAAATGCCCTAGTTAATGATGCTGTAATCCAGATTATCA 63
    |||
QY 364 ATCATCTTTGAGAAAGCTTCGAGGCCAATATCGCGTCATTGTGAATGTTTAA 423
    |||
DB 64 ATTACCTTTCAAAATGATTAGTAAAGCAATATCGCTATATTGTTAATGTCATGA 123
    |||
QY 424 GACCATATGATGTCATGAGGCGGCGGCGGTCGATGAAATGCGAAGCCGTTACCGCTACA 483
    |||
DB 124 GACCATATGATGTCATGAGGCGGCGGCGGTCGATGAAATGCGAAGCCGTTACCGAACA 183
    |||
QY 484 ATTCCTTATATGCGCATCTTCTGATTCAGATGATGATATACCGAGTTCTTAAACAA 543
    |||
DB 184 ATTCATATATACGGAAATTTAGTTGTAATGAAAGATTAATCTATCAATTTCTTGCAAG 243
    |||
QY 544 AAACGAAAGAGCAAAACACAAAAGTCATCTGCTGTAATCTCAAAAATTAACAGATGAG 603
    |||
DB 244 GAACGTAAAAAGCGTAAATTCAGAACTCTGTTGTGATTAAGACGTATACAGAAATCA 303
    |||
QY 604 TATTTACGTAATTTTGAATACATGATATCCCTGATTAACGCTTCTGCGCTGGGTGTG 663
    |||
DB 304 TATTTACGGAAGTTGATTTATTTAGTATTTCTGTATATGATGATGTTGTTAGAAATA 363
    |||
QY 664 GCTAAGACCTCGCATGAGAAAGAACACATTTTAAGGAAATGCTGAATGCCCGCCA 723
    |||
DB 364 GCGCAAGAGTTGGTATGATGAAGAACTGCATTTACAAAGGTATGTTAAATGACACCAACC 423
    |||
QY 724 GATCCGGAGCAATGAGAAATCTTCGCGTATCAAGTCCGAGCGAGCCGCGACATTTGTT 783
    |||
DB 424 GATCCAGGTGCTGTTAGATTAATTTCCATGCAAAATCCGCAAAAAAATGATTTGTT 483
    |||
QY 784 AATGGGTTTGGCGGAAAGCAGCGCTTCTTCTAATTTGAATATATGAAACGTGTAAAGAA 843
    |||
DB 484 AATGCATTCGCTGCTAATGAAACCGCAGCTACAAAAGGATTTTAAATGAAGTGAATCA 543
    |||
QY 844 ATCGGTTAACCGGATGATCCGATCATCATGAACTGCGCGGCGAGACCGGTGCAT 903
    |||
DB 544 TATTAATTAATCATGATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 603
    |||
QY 904 CGGACACGCAATTCGCAATATGACGATTTGCTTATTTATGAAGCAAGCACTGATCTTA 963
    |||
DB 604 AGAACACAACTCTTTGTTGATTAATCTTTTAGGTGAAGTCAATGATGTTCTCATTTGT 663
    |||
QY 964 ATCGGTGAACCAACGAACGAT 986
    |||
DB 664 ACAGGAAAAAGTACAAATGAT 686
    |||
```

RESULT 15
ABK75876
_id ABK75876 standard; DNA; 309 BP.

```
XX ABK75876;
AC 13-AUG-2002 (first entry)
DT Bacillus licheniformis genomic sequence tag (GST) #3167.
DE Bacillus licheniformis genomic sequence tag (GST) #3167.
KW Differential gene expression; genomic sequenced tag; GST;
KM altered culture condition; environmental stress;
KW physiological provocation; de.
XX Bacillus licheniformis.
OS Bacillus licheniformis.
PN WO200229113-A2.
XX 11-APR-2002.
PD 05-OCT-2001; 2001WO-US031437.
XX 06-OCT-2000; 2000US-00680598.
XX 27-MAR-2001; 2001US-0279526P.
PR (NOVO ) NOVOTYMES BIOTECH INC.
PA (NOVO ) NOVOTYMES AS.
XX Berka R, Clausen IG;
PI WPI: 2002-416684/44.
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX Claim 4; SEQ ID NO 3167; 200pp; English.
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX relative gene expression by an observed hybridisation reporter signal of
XX a spot in the array. The method is useful for measuring the expression of
XX genes in a first Bacillus cell relative to expression of the same genes
XX in one or more second Bacillus cells. The method is useful for monitoring
XX global expression of several genes from a Bacillus cell, discovering new
XX genes, identifying possible functions of unknown open reading frames and
XX monitoring gene copy number variation and stability. Monitoring changes
XX in expression of genes may be used to provide a representation of the way
XX in which Bacillus cells adapt to changes in culture conditions,
XX environmental stress or other physiological provocation. Extensive follow
XX up characterisation is unnecessary, when one spot on an array equals one
XX gene or one open reading frame, since sequence information is available.
XX This sequence represents a genomic sequence tag (GST) used in the method
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 309 BP; 100 A; 62 C; 76 G; 71 T; 0 U; 0 Other;

Query Match 15.3%; Score 181; DB 6; Length 309;  
Best Local Similarity 74.1%; Pred. No. 1.3e-40;  
Matches 229; Conservative 0; Mismatches 80; Indels 0; Gaps 0;



```
QY 873 CATATGAACCTGCGCGGAGACCGTGTGATCGAGACACGCAATTCGCAATGACGAT 932
 |||
DB 1 CCTATTATTAATGCGCGCGGAGACCGTGTGATCGAGACACGCAATTCGCAATGACGAT 60
 |||
QY 933 GCCTTATATGAAGCAAGTGAATGATCTTAATGCGTGAAGCAACAGACGATCGTAA 992
 |||
DB 61 TCCTTATATGAAGCAAGTGAATGATCTTAATGCGTGAAGCAACAGACGATCGTAA 120
 |||
QY 993 AGCCTATGAAGCAAGCAAAATTCCTGCGAGCAAACTGCATGACCTAGATTAAGTCAAC 1052
 |||
```


```


Db	121	AGCATATGAGCAGGCAAAATTCTGCGACAACTGTTGATTTTGAGCACAATCAAC	180
Qy	1053	AGATGAATTTATGAAATTGTTAAGAAAAAGAAATGCAACCGTGTATATATATGCGTCGG	1112
Db	181	GGAAGAAATCATGTCATGCTGAAAAACAAGCTTGAGGGCCGCGTTATTACGAGTCGG	240
Qy	1113	CAATATTCATGTCGCCGACAGACCTTTAATTGAAAAATCCAGAAATACAGGTAAAGCA	1172
Db	241	AAATATCCAGGAGCGAGCCTCTCTCATTGAAAAAATGCAAGATTACAGATTAAAGCA	300
Qy	1173	GTCGTTAG	1181
Db	301	GTCGTTAG	309

Search completed: February 27, 2006, 07:05:15
 Job time : 756.404 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28, Search time 1294.32 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182
Sequence: 1 atggcgctggttactctatc.....agtaagcagctcgaacg 1182

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA_Main:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1182	100.0	1182	8 US-10-484-605-1	Sequence 1, Appl1
2	1182	100.0	1182	9 US-10-789-164-1	Sequence 1, Appl1
3	1182	100.0	6536	8 US-10-484-605-6	Sequence 6, Appl1
4	548.4	46.4	819	3 US-09-974-300-2694	Sequence 2694, Ap
5	338.6	28.6	1164	7 US-10-724-972A-814	Sequence 814, Ap
6	313.2	26.5	1212	9 US-10-501-282-53	Sequence 53, Appl
7	313.2	26.5	1754382	9 US-10-501-282-6651	Sequence 6651, Ap
8	181	3.8	1338	3 US-09-974-300-3167	Sequence 3167, Ap
9	44.4	3.6	600	7 US-10-282-122A-10762	Sequence 10762, A
10	42.6	3.6	454	3 US-09-814-353-17917	Sequence 17917, A
11	42.2	3.6	12507	6 US-10-311-455-272	Sequence 272, App
12	42	3.6	3673778	6 US-10-311-455-2278	Sequence 2278, Ap
13	41.4	3.5	6000	6 US-10-311-455-1788	Sequence 1788, Ap
14	40	3.4	7312	7 US-10-221-613-312	Sequence 312, App
15	40	3.4	7312	7 US-10-221-613-312	Sequence 312, App
16	39.8	3.4	11944	6 US-10-311-455-2159	Sequence 2159, Ap
17	39.6	3.4	1047	6 US-10-724-972A-3451	Sequence 3451, Ap
18	39.6	3.4	1546	6 US-10-017-161-2325	Sequence 2325, Ap
19	39.6	3.4	1546	6 US-10-292-798-1971	Sequence 1971, Ap
20	39.6	3.4	519	5 US-10-027-632-68820	Sequence 68820, A
21	39.2	3.3	519	5 US-10-027-632-68821	Sequence 68821, A
22	39.2	3.3	519	5 US-10-027-632-68821	Sequence 68821, A
23	39.2	3.3	519	5 US-10-027-632-68822	Sequence 68822, A

C	24	39.2	3.3	519	5	US-10-027-632-298134	Sequence 298134,
C	25	39.2	3.3	519	5	US-10-027-632-298135	Sequence 298135,
C	26	39.2	3.3	519	6	US-10-027-632-68820	Sequence 68820, A
C	27	39.2	3.3	519	6	US-10-027-632-68821	Sequence 68821, A
C	28	39.2	3.3	519	6	US-10-027-632-68822	Sequence 68822, A
C	29	39.2	3.3	519	6	US-10-027-632-298134	Sequence 298134,
C	30	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	31	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	32	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	33	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	34	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	35	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	36	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	37	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	38	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	39	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	40	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	41	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	42	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	43	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	44	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	45	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,

ALIGNMENTS

RESULT 1
US-10-484-605-1
Sequence 1, Application US/10484605
Publication No. US20040253704A1
GENERAL INFORMATION:
APPLICANT: Sung, Moon-Hee
APPLICANT: Hong, Seung-Pyo
APPLICANT: Lee, Jong-Su
APPLICANT: Jung, Chang-Min
APPLICANT: Kim, Chul-Doong
APPLICANT: Soda, Kenji
APPLICANT: Ashbaugh, Makoto
TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING P88BC, THE GENE CODING
TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
TITLE OR INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
FILE REFERENCE: 4240-101
CURRENT APPLICATION NUMBER: US/10/484,605
CURRENT FILING DATE: 2004-01-20
PRIOR APPLICATION NUMBER: PCT/KR02/01522
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: KR 2001-48373
PRIOR FILING DATE: 2001-08-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 1182
TYPE: DNA
ORGANISM: Bacillus subtilis
US-10-484-605-1

Query Match 100.0%; Score 1182; DB 8; Length 1182;
Best Local Similarity 100.0%; Pred. No. 7.6e-312; Indels 0; Gaps 0;

Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGCGCTGTTACTCTATTATAGCTCTGTCATCTGTCATCGGAATATTAGAAAA	60
DB	1	ATGGCGCTGTTACTCTATTATAGCTCTGTCATCTGTCATCGGAATATTAGAAAA	60
QY	61	CGACGACATCAGAAAAATTGATGCGCTCCCTGTTCCGGTGAATATTACGGCATCCGC	120
DB	61	CGACGACATCAGAAAAATTGATGCGCTCCCTGTTCCGGTGAATATTACGGCATCCGC	120
QY	121	GGAAAAATCGACTGTGCAAGGCTGACACCGGAATATTATAGAACCGGTTTCAAGACT	180
DB	121	GGAAAAATCGACTGTGCAAGGCTGACACCGGAATATTATAGAACCGGTTTCAAGACT	180

181 GTTGGAAAAACAAGAGAGATGACAGAAATTTACTGGGACACACCGGAGGAAAAG 240
181 GTTGGAAAAACAAGAGAGATGACAGAAATTTACTGGGACACACCGGAGGAAAAG 240
241 CCGATTAAACGGAACCTCAGGGGCGGAATTCGGAGAGCAAAAAGAGTCAAGAGAA 300
241 CCGATTAAACGGAACCTCAGGGGCGGAATTCGGAGAGCAAAAAGAGTCAAGAGAA 300
301 ACAGTAGAAAGGGGCTAAGCGCATGTCAGTGAATGTCAGTGGCTGTTAACCCGATTTAT 360
301 ACAGTAGAAAGGGGCTAAGCGCATGTCAGTGAATGTCAGTGGCTGTTAACCCGATTTAT 360
361 CAAATCATCTTTCAGAGAACTTCGAGGCGCAATATCGCGCTCATTTGTAATGTTTTTA 420
361 CAAATCATCTTTCAGAGAACTTCGAGGCGCAATATCGCGCTCATTTGTAATGTTTTTA 420
421 GAAAGCATATGATGTCATGGGGCGGACGCTTGATGAATTTGCAAGAGCGTTTACCGCT 480
421 GAAAGCATATGATGTCATGGGGCGGACGCTTGATGAATTTGCAAGAGCGTTTACCGCT 480
481 ACAATTCCTTAAATGGCATCTTGTCACTAAGATAGTAATACCGAGTTCTTTTAA 540
481 ACAATTCCTTAAATGGCATCTTGTCACTAAGATAGTAATACCGAGTTCTTTTAA 540
541 CAAAAAGCAAAAGAGAGAGCAAAAGTCAATTCATTCGATTAATCTCAAAAATTAACAGAT 600
541 CAAAAAGCAAAAGAGAGAGCAAAAGTCAATTCATTCGATTAATCTCAAAAATTAACAGAT 600
601 GAGTATTTAAGTAAATTTGAAATACATGATATTCCTGATTAACGCTTCTGCGCTGGGT 660
601 GAGTATTTAAGTAAATTTGAAATACATGATATTCCTGATTAACGCTTCTGCGCTGGGT 660
661 GTGGCTCAAGACCTCGGATGAGAGAAACAGATTTAAAGGGAATGCTGAATGCGCG 720
661 GTGGCTCAAGACCTCGGATGAGAGAAACAGATTTAAAGGGAATGCTGAATGCGCG 720
721 CCAAGTCCGGAGAGCAATGAGAAATTTCTCGCTGATCAGTCCGAGAGAGCTGGGCACTTT 780
721 CCAAGTCCGGAGAGCAATGAGAAATTTCTCGCTGATCAGTCCGAGAGAGCTGGGCACTTT 780
781 GTTAAATGGGTTTGGCGCAACGACGCTTCTTCTAATTTGAATATATGAAACGTTTAA 840
781 GTTAAATGGGTTTGGCGCAACGACGCTTCTTCTAATTTGAATATATGAAACGTTTAA 840
841 GAAATCGGTTTCCGAGAGAGAGATGAGAAATTTCTCGCTGATCAGTCCGAGAGAGCTGGG 900
841 GAAATCGGTTTCCGAGAGAGAGATGAGAAATTTCTCGCTGATCAGTCCGAGAGAGCTGGG 900
901 GATCGAGACAGCAATTCGCAAAATGAGATTCCTTATATTTGAAGCAAGTGAATGATC 960
901 GATCGAGACAGCAATTCGCAAAATGAGATTCCTTATATTTGAAGCAAGTGAATGATC 960
961 TTAATCGGTTTAAACAGAAACGATTCGTAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
961 TTAATCGGTTTAAACAGAAACGATTCGTAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
1021 GACAAACCTGACATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGA 1080
1021 GACAAACCTGACATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGA 1080
1081 AGAATGACAAACCTGATATATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGA 1140
1081 AGAATGACAAACCTGATATATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGA 1140
1141 ATTGAAAAAATTCAGCAATTAAGGTTAAAGAGCTGTAAGC 1182
1141 ATTGAAAAAATTCAGCAATTAAGGTTAAAGAGCTGTAAGC 1182

RESULT 2
US-10-789-164-1
; Sequence 1, Application US/10789164
; Publication No. US20050191720A1

GENERAL INFORMATION:
; APPLICANT: Sun, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; TITLE OF INVENTION: ENCODING POLY-GLUTAMATE SYNTHETASE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: KopelecIn 1.71
; SEQ ID NO 1
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-789-164-1
Query Match 100.0%; Score 1182; DB 9; Length 1182;
Best Local Similarity 100.0%; Pred. No. 7,6e-312; Mismatches 0; Gaps 0;
Matches 1182; Conservative 0; Indels 0; Gaps 0;
1 ATGGGCTGTTACTCATTTATAGCTGCTGTCATCTGTCATCGGATATTTAGAAAA 60
1 ATGGGCTGTTACTCATTTATAGCTGCTGTCATCTGTCATCGGATATTTAGAAAA 60
61 CGACGACATCAAAAAACATTGATGCCCTCCCTGTTGGGTTGAATTTAACGCAATCCG 120
61 CGACGACATCAAAAAACATTGATGCCCTCCCTGTTGGGTTGAATTTAACGCAATCCG 120
121 GGAATATGACTGTCAGCAAGGCTGCAACCGGAATTTAATAGAGCCGGTTACAGACT 180
121 GGAATATGACTGTCAGCAAGGCTGCAACCGGAATTTAATAGAGCCGGTTACAGACT 180
181 GTTGGAAAAACAAGAGAGATGACAGAAATTTACTGGGACACACCGGAGGAAAAG 240
181 GTTGGAAAAACAAGAGAGATGACAGAAATTTACTGGGACACACCGGAGGAAAAG 240
241 CCGATTAAACGGAACCTCAGGGGCGGAATTCGGAGAGCAAAAAGAGTCAAGAGAA 300
241 CCGATTAAACGGAACCTCAGGGGCGGAATTCGGAGAGCAAAAAGAGTCAAGAGAA 300
301 CCAATTCATCTTTCAGAGAACTTCGAGGCGCAATATCGCGCTCATTTGTAATGTTTTA 420
301 CCAATTCATCTTTCAGAGAACTTCGAGGCGCAATATCGCGCTCATTTGTAATGTTTTA 420
421 GAAAGCATATGATGTCATGGGGCGGACGCTTGATGAATTTGCAAGAGCGTTTACCGCT 480
421 GAAAGCATATGATGTCATGGGGCGGACGCTTGATGAATTTGCAAGAGCGTTTACCGCT 480
481 ACAATTCCTTAAATGGCATCTTGTCACTAAGATAGTAATACCGAGTTCTTTTAA 540
481 ACAATTCCTTAAATGGCATCTTGTCACTAAGATAGTAATACCGAGTTCTTTTAA 540
541 CAAAAAGCAAAAGAGAGAGCAAAAGTCAATTCATTCGATTAATCTCAAAAATTAACAGAT 600
541 CAAAAAGCAAAAGAGAGAGCAAAAGTCAATTCATTCGATTAATCTCAAAAATTAACAGAT 600
601 GAGTATTTAAGTAAATTTGAAATACATGATATTCCTGATTAACGCTTCTGCGCTGGGT 660
601 GAGTATTTAAGTAAATTTGAAATACATGATATTCCTGATTAACGCTTCTGCGCTGGGT 660
661 GTGGCTCAAGACCTCGGATGAGAGAAACAGATTTAAAGGGAATGCTGAATGCGCG 720
661 GTGGCTCAAGACCTCGGATGAGAGAAACAGATTTAAAGGGAATGCTGAATGCGCG 720
721 CCAAGTCCGGAGAGCAATGAGAAATTTCTCGCTGATCAGTCCGAGAGAGCTGGGCACTTT 780
721 CCAAGTCCGGAGAGCAATGAGAAATTTCTCGCTGATCAGTCCGAGAGAGCTGGGCACTTT 780
781 GTTAAATGGGTTTGGCGCAACGACGCTTCTTCTAATTTGAATATATGAAACGTTTAA 840
781 GTTAAATGGGTTTGGCGCAACGACGCTTCTTCTAATTTGAATATATGAAACGTTTAA 840

Db	781	GTTAATGGGTTTGGCCGCAAAAGACGCTTCTTCACTTGAATATATAGAAACGGTAA	840
Qy	841	GAATTCGGTATCCCGACCGATGATCCGATCATCATGAACTCCCGCGACACCGTGT	900
Db	841	GAATTCGGTATCCCGACCGATGATCCGATCATCATGAACTCCCGCGACACCGTGT	900
Qy	901	GATGGGACACAGCAATTCCGCAATGAGCTATGTCCTATATGGAAGCAAGAACTATC	960
Db	901	GATGGGACACAGCAATTCCGCAATGAGCTATGTCCTATATGGAAGCAAGAACTATC	960
Qy	961	TTAATCGGTGAAACAACAGAACCGATCGTAAAGCCTATGAAGAGCAAAATTCCTG	1020
Db	961	TTAATCGGTGAAACAACAGAACCGATCGTAAAGCCTATGAAGAGCAAAATTCCTG	1020
Qy	1021	GACAAACTGCATGACTTGAATAGTCAACAGTGAATTTATGAAATTTGTTAAAGAA	1080
Db	1021	GACAAACTGCATGACTTGAATAGTCAACAGTGAATTTATGAAATTTGTTAAAGAA	1080
Qy	1081	AGATTCGCAACCCGTCATATATGTCGGCTCGGCAATATTCATGGTGCCGACAGGCTT	1140
Db	1081	AGATTCGCAACCCGTCATATATGTCGGCTCGGCAATATTCATGGTGCCGACAGGCTT	1140
Qy	1141	ATTGAAAAATTCACGAAATACAGGTAAAGACGCTCGTAAGC	1182
Db	1141	ATTGAAAAATTCACGAAATACAGGTAAAGACGCTCGTAAGC	1182

RESULT 3

US-10-484-605-6

Sequence 6, Application US/10484605

Publication No. US20040253704A1

GENERAL INFORMATION:

APPLICANT: Sung, Moon-Hee

APPLICANT: Hong, Seung-rye

APPLICANT: Lee, Jong-su

APPLICANT: Jung, Chang-Min

APPLICANT: KLM, CHILD-DOING

APPLICANT: Ashiuchi, Makoto

APPLICANT: ABILUCH, MAXIM
TITLE OF INVENTION: SURFACE

TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSING AND PURIFYING THE SAME

TITLE OF INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR

FILE REFERENCE: 4240-101

CURRENT APPLICATION NUMBER

CURRENT FILING DATE: 2004-01-20

PRIOR APPLICATION NUMBER: PCT/KR02/015222

PRIOR FILING DATE: 2002-08-09

; PRIOR APPLICATION NUMBER: KR 2001-48373

PRIOR FILING DATE: 2001-08-10 ;

NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 6

; LENGTH: 6536

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Synthetic construct

US-10-484-605-6

Query Match	Score	DB	Length
100.0%	1182	8	6536

Best Local Similarity 100.0%; Pred. No. 1.9e-311;

Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY _ AAGGGCTGGTTCATTTATPAGCCTGTGCTGCATACCTGATCGAATATTAGAAAA 60
Db 224 ATGGGCTGGTTCATTTATAGCCTGTGCTGCATACCTGATCGAATATTAGAAAA 283
QY 61 CGACGACATCAGAAAAACATTTGATGCCCTCCCTGTTGGGGTGAAATTTAAACGCATCCGC 120
Db 284 CGACGACATCAGAAAAACATTTGATGCCCTCCCTGTTGGGGTGAAATTTAAACGCATCCGC 343
QY 121 GGAATAATCGACTGTGCAAGGCTGACACCGGAATTTAATTAGAACCGGTTTCAAGACT 180

Db	344	CGAATAATCGACTGTGTGACAAGCGTGCACACCGAATATTAAATAGAAAGCCGTTTACAAAGCT	403
Oy	181	GTTGGAAAAACACACAGAAACAGATCTCAAGATGATTTACTGGGACACACCGSAGAGAAAG	240
Db	404	GTTGGAAAAAACACACAGAACAGATCTCAAGATGATTTACTGGGACACACCGSAGAGAAAG	463
Oy	241	CGGATTTAAACGGAAACCTCAGAGGGCCGAATATCGGAGAGCAAAAAGAGTATGAGAGAA	300
Db	464	CCGATTTAAACGGAAACCTCAGAGGGCCGAATATCGGAGAGCAAAAAGAGTATGAGAGAA	523
Oy	301	ACAGTAGAAAAAGGGCTTAACCGCATTTGTCAGTGAATGCATGGCTGTTAACCCAGATTAT	360
Db	524	ACAGTAGAAAAAGGGGCTTAACCGCATTTGTCAGTGAATGCATGGCTGTTAACCCAGATTAT	583
Oy	361	CAAAATCATCTTTACAGAAAGAACTTTGTGACAGCGCAATATGGGCGTCACTTGTAATGTTTTA	420
Db	584	CAAAATCATCTTTACAGAAAGAACTTTGTGACAGCGCAAAATATGGGCGTCACTTGTAATGTTTTA	644
Oy	421	GAAGACCAATATGAGATTCATATGGGGCGACGCTTGATGAAATTCAGAAAGCGTTTACCGCT	480
Db	644	GAAGACCAATATGAGATTCATATGGGGCGACGCTTGATGAAATTCAGAAAGCGTTTACCGCT	703
Oy	481	ACAATTCCTTAATATGAGCCATCTTGTCAATTACAGATAGTGAATATACCGAATTTCTTTAAA	540
Db	704	ACAATTCCTTAATATGAGCCATCTTGTCAATTACAGATAGTGAATATACCGAATTTCTTTAAA	763

20

Db 644 GAAG

QY	481	ACAAATTCCTTAATGAGCCATCTGTGCATTAACAGTAGAATATACGAGTCTTTAAA	54
Db	704	ACAATTCCTTAATATGGCCATCTGTGCATTAACAGTAGAATATACGAGTCTTTAAA	76
QY	541	CAAAAAGCAAAAGACGAAACACAAAAGTCATCATTTGCGATACTCAAAAATTACAGAT	60
Db	764	CAAAAAGCAAAAGACGAAACACAAAAGTCATCATTTGCGATACTCAAAAATTACAGAT	82
QY	601	GAGATTTAGTAATTTTGAATACATNGEATTCCTGATTAACGCTTCCTGAGCGTGGGT	66
Db	824	GAGATTTAGTAATTTTGAATACATNGEATTCCTGATTAACGCTTCCTGAGCGTGGGT	88
QY	661	GTGGCTCAACCACTCGGCATTTGACGAAAGAACGACATTTAAGGAGATCTGAATCCGCG	72
Db	884	GTGGCTCAACCACTCGGCATTTGACGAAAGAACGACATTTAAGGAGATCTGAATCCGCG	94
QY	721	CCAGATCCGGAGACCAATGAGAAATCTTCCGCTGATCAAGTCCGAGCGACCTTGGCACTTT	78
Db	944	CCAGATCCGGAGACCAATGAGAAATCTTCCGCTGATCAAGTCCGAGCGACCTTGGCACTTT	100
QY	781	GTTAATGGGTTTCCCGCAAAACGACGCTTCTTCTACTTTGAATATATGCAAAACGTGTAAA	84
Db	1004	GTTAATGGGTTTCCCGCAAAACGACGCTTCTTCTACTTTGAATATATGCAAAACGTGTAAA	100
QY	841	GAAATCGGTTAACCCGACCGATGATCCGATCATCATGAAATGCGCGGACGACCGGTGC	90
Db	1064	GAAATCGGTTAACCCGACCGATGATCCGATCATCATGAAATGCGCGGACGACCGGTGC	110
QY	901	GATCGGACACAGCAATTCGCAATGAGAGTATGCTTATATGGAAGCAAGTGAATCTGATC	96
Db	1124	GATCGGACACAGCAATTCGCAATGAGAGTATGCTTATATGGAAGCAAGTGAATCTGATC	112
QY	961	TTAATCGGTGAAACACAGAACCGATCGTAAAGCCTATGAAGAGCAAAATTCCTGCA	102
Db	1184	TTAATCGGTGAAACACAGAACCGATCGTAAAGCCTATGAAGAGCAAAATTCCTGCA	118
QY	1021	GACAAACTGCATGACCTAGAGTAAAGTCAACAGATGAATATATGAAATTTGTTAAAGAAA	108
Db	1244	GACAAACTGCATGACCTAGAGTAAAGTCAACAGATGAATATATGAAATTTGTTAAAGAAA	124
QY	1081	AGAAATGCAACAACCGTGTCAATATATGGCGTGGGCAATATTCATATGTCGCCGACAGCTTTA	114
Db	1304	AGAAATGCAACAACCGTGTCAATATATGGCGTGGGCAATATTCATATGTCGCCGACAGCTTTA	130
QY	1141	ATTGAAAAAATCCACGAATCAAGGTAAAGCAGCTGTGAAGC 1182	
Db	1364	ATTGAAAAAATCCACGAATCAAGGTAAAGCAGCTGTGAAGC 1405	

Db 1065 GAAATCCTTGGAGATATTACATGACCTTAATGACTTATGCTTGTGTAAGTTGCTT 1124
QY 1104 TGGCGTCGGCAATATTTCATGTGCGCGCAGAGCCTTTAATTGAAAAAATCCAGAAATACA 1163
Db 1125 GGGATTGGAATATTCATGTGTGCGTAAGATTTTTTGTCTTACCTTGAATGAAAAATA 1184
QY 1164 GGTAAAGCAG 1173
Db 1185 CTTGTGCGAG 1194

RESULT 7
US-10-501-282-6651/c
; Sequence 6651, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOTOCOCUS OTTIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT FILING DATE: 2004-07-09
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6651
; LENGTH: 1754382
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
US-10-501-282-6651

Query Match 26.5%; Score 313.2; DB 9; Length 1754382;
Best Local Similarity 55.1%; Pred. No. 7.7e-73;
Matches 634; Conservative 0; Mismatches 513; Indels 3; Gaps 1;

QY 27 TGGCTGCTACTGTCATCGCAATATTAGAAAAAGAGACATCAAGAAAAATGATGTC 86
Db 23405 TGGCTCTCTTATTGGCTTAGGCTTATAGAAAGCAGCTTAACGAGAAATATTAAAA 23346
QY 87 CCTCCCTGTTCCGGTGAATATTACGGCATCCGCGAAATCGACTGTGACAAAGCTGAC 146
Db 23345 AATGACCTAGCATTAAATTAATGCGACCGAGGCAAGACACAGTACTGTTGGT 23286
QY 147 AACCGAATATTAAATGAAGCCGGTTACAAAGCTGTGAAAAACAACAGAAACAGATGC 206
Db 23285 TACTGGGGGTCTACAAAGCGCTGCTACCAAGTTATTGAAAAACAACAGAACTGAGTC 23226
QY 207 AAGAAATGATTTCTGCGACACACCGAGAGAAAAAGCGAATTAACGAAACCTCAGGGGCC 266
Db 23225 ACGCATTAATTACTGGAGACAGGCAAGAAAGAGAGAGATCGAAAGCCCTCCATCTGGACC 23166
QY 267 GAATATCGGAGAGCAAAAAAGAGTCATGAGAGAAACAGTAGAAAGAGGGGCTAACGCCAT 326
Db 23165 TAAATTTCTGAAACAAATATCAGTTATTGAAAAAGCAAAAAAGAGGGGCTGATGCCCT 23106
QY 327 TGTCAATGATGATGATGCTGTTTAACCCAGATTATCAATATCATCTTTACAGAAATCTCT 386
Db 23105 TGTATTAGAAATGATGAGCAGTTAATCAGAGTACCAAGATGTTATTCTAATGAATGTT 23046
QY 387 GGAGGCCAATATCGGGCTATGTGAATGTTTAAAGAACCTATGATGATGATGAGGCC 446
Db 23045 TCAATGAATATTAACGACCATGTTAACTTTGAAAGCAATATGAGATATCTAGGCTC 22986
QY 447 GACGCTTGATGAATATGCAAGCGTTTACCGCTACAAATTCCTTAAATGAGCATCTTGT 506

Db 22385 AACAAATGATCAATTCCTGCGCTATGCGCAAGACCATCCCAAAATGCGCTTGTGT 22926
QY 507 CATTAAGATATGATATATCCAGATCTTTTAAACAAAAAGCAAAAGCAACAA 566
Db 22325 CATCTGAGATGATACATTAAGACTTTTATTTAGAGTACCCAGAGAGAAATTCGCC 22866
QY 567 AGTCATGATTCGTATTAATCAAAAATTAAGATGATGATTTAATTTGAAATCAT 626
Db 22865 GGTGAAGTTATATACGACACAGCGCTTATGCAAGATTAATGATGCTTGTGCTATCT 22806
QY 627 GGTATTCCTGATATACGCTTCTGCGCGCTGGGTGTGCGCTCAAGCATTCGGCATTTGACG 686
Db 22805 TGTCTCCCAAAACATCTAGCTGTTCCGCTGGCTATGCTGAAGAACTTCACGTTGACG 22746
QY 687 AGAAACGATTTTAAGGAAATGCTGAATGCGCGCCAGATCCGGGACCAATGAGAAATCT 746
Db 22745 CGATACGCAATTAAGATATGCTTCAAGCTGACCTGATTCGGGCAATTTGCAATGTT 22686
QY 747 TCCGCTGATCAGTCCGAGCGAGCCTGGGCACTTT--GTTAATGGCTTGGCCGAAACGA 803
Db 22685 TGAATAGGGGAGGCTATTTGACAGAAAGTTATTTCTAGCTTATGCTTGTGCTAATGA 22626
QY 804 CGCTTCTTCTACTTGAATATATGAAAACGTGTAAGAAATCGGTTAACCGACCATGA 863
Db 22625 ACCCATCTTACCATGCTGATCTATGAAAGCTCAAGAAAGTGGCTTAATAGGTGAAG 22566
QY 864 TCCGATATCATCATGAACTGCGCGGCGAGCCGCTGATGCGACAGCAATTCGCAAA 923
Db 22565 CCTATTATTTTGAATGATGCTGCGAGAAAGCAGTACCGGCTTAATTTATTCACCA 22506
QY 924 TGACGATTCCTTATATTGAAAGCAATGATCTTATTCGTTGAAACACAGAAC 983
Db 22505 AGACTTATTTGCTAAATTAAGATGCGAGGCTATGCGCATTTGGAACAAATATCTGGCAT 22446
QY 984 GATGTAAGCCATATGAGAGGCAAAATTCCTGACAGCAAAATCGATGACCTAGAGTA 1043
Db 22445 TATTGAGAGGCTTAAATACGATGAGATCAAGCTCAAGGATTAATTTATCTTTGAGG 22386
QY 1044 TAATGCAACAGATTAATTAATGATTTGTTAAAGAAAGATGCAACACGCTCATATA 1103
Db 22385 GAAATCTTGAAGATTAATTAATGATGACCTTAATGATTTGCTGTGAAATGTTCTT 22326
QY 1104 TGGCGTCGGCAATATTTCATGTGCGGAGCCCTTAATGAAAAATCCAGAAATCAA 1163
Db 22325 GGGTATTGGAATATTAATGATGCTGCGTGAAGATTTTTTGTCTTACCTGATGAAAAATA 22266
QY 1164 GGTAAAGCAG 1173
Db 22265 CTTGTGCGAG 22256

RESULT 8
US-09-974-300-3167
; Sequence 3167, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3167
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Bacillus licheniformis

LENGTH: 6000
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2278

Query Match 3.5%; Score 41; DB 6; Length 6000;
Best Local Similarity 48.1%; Pred. No. 8.5; Indels 0; Gaps 0;
Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 428 ATATGGATGTCATGGGCGCAGCCCTGATGAAATTCAGAGAGGTTACCGCTACATTC 487
DB 5183 ATTTGGAGCTTGAGGTGAGAGATTATTGAAATTTAGAGTTGATGAGTTGTGTGA 5242
QY 488 CTTAATATGCGCATCTTGTCATACAGATAGTAATACCGAGTTCTTTAAACAAAAG 547
DB 5243 GTTATGATGCGGTATTATTATTGAGTAAAGTGAATTTGTTTAAAAAT 5302
QY 548 CAAAGACGAAACACAAAGTCATCTGATACCAAAATTAACAGATGATTT 607
DB 5303 AAAAAATAAAAATATTAATAATTTGTAATTAATAATTAATAATAAATTAAT 5362
QY 608 TACGTAATTTGATCATGCTTCCCGATTAACGCTCTCGCGCTGGGTGCTC 667
DB 5363 ATAAAAATTATTATGATTAATGTTTTTTATTAATGTTGAGATTTTGGGTGAGATTA 5422
QY 668 A 668
DB 5423 A 5423

RESULT 15

US-10-311-455-1788/C
Sequence 1788, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1788
LENGTH: 7312
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1788

Query Match 3.4%; Score 40; DB 6; Length 7312;
Best Local Similarity 53.1%; Pred. No. 18;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 958 ATCTTAATCGGTAAACAAGAACCGATGCTAAAGCTATGAGAAGCAAAATTCCT 1017
DB 4098 AACTTAATCAACAAATAAAACCAACCAACCAACCAACCAACCAACCAACCAAC 4039
QY 1018 GCAGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1077
DB 4038 GCAGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3979
QY 1078 AAAAGATGACAAACGTCATATATGCGCTCGCAATA 1117

DB 3978 TAACGCTCTCCACAACTCAAAAAAAGATTAACAAA 3939

Search completed: February 27, 2006, 07:58:23
Job time : 1304.32 secs

THIS PAGE BLANK (USPTO)

Qy	181	GTGGGAAAAACAACAGGAAACGATGCGAAGATGATTTTACTGGGACACACCGGAGGAAAAG	240
Dp	181	GTGGGAAAAACAACAGGAAACGATGCGAAGATGATTTTACTGGGACACACCGGAGGAAAAG	240
Qy	241	CCGATTTAAACGGAAACCTCAGGGGGCCGATATCGGAGAGCAAAAAGAAGTCATGAGAA	300
Dp	241	CCGATTTAAACGGAAACCTCAGGGGGCCGATATCGGAGAGCAAAAAGAAGTCATGAGAA	300
Qy	301	ACAGTAGAAAAGGGGGCTTACCGCATTTGTCACTGATATGCAATGGCTTTTAAACCAGATTTAT	360
Dp	301	ACAGTAGAAAAGGGGGCTTACCGCATTTGTCACTGATATGCAATGGCTTTTAAACCAGATTTAT	360
Qy	361	CAAAATCATCTTTACAGAAAGAACTTCTCGAGGSCCAATATCGCGCTCATTTGTAAATGTTTTA	420
Dp	361	CAAAATCATCTTTACAGAAAGAACTTCTCGAGGSCCAATATCGCGCTCATTTGTAAATGTTTTA	420
Qy	421	GAAGACCATATGATGATGTCATGGGGCCGACGCTTGATGAAATTTGCAAGAACGTTTAAACCGCT	480
Dp	421	GAAGACCATATGATGATGTCATGGGGCCGACGCTTGATGAAATTTGCAAGAACGTTTAAACCGCT	480
Qy	481	ACAATTCCTTTAATATGGGCATCTTGTCATTAACAGTAGTGAATATATACCGAATTCCTTTAAA	540
Dp	481	ACAATTCCTTTAATATGGGCATCTTGTCATTAACAGTAGTGAATATATACCGAATTCCTTTAAA	540
Qy	541	CAAAAAGCAAAAGAACGAAACACAAAAGTCATCATTTGCTGATTAATCTCAAAAATTTACAGAT	600
Dp	541	CAAAAAGCAAAAGAACGAAACACAAAAGTCATCATTTGCTGATTAATCTCAAAAATTTACAGAT	600
Qy	601	GAGTATTTTACGTAATTTTGAATACATGATGATATCCCTGATAACGCTTCTTGCGCCCTGGGT	660
Dp	601	GAGTATTTTACGTAATTTTGAATACATGATGATATCCCTGATAACGCTTCTTGCGCCCTGGGT	660
Qy	661	GTGGCTCAAGACACTCCGCAATTGACGAAAGAACAGATTTAAGGAAATGCTGAATAGCGCGG	720
Dp	661	GTGGCTCAAGACACTCCGCAATTGACGAAAGAACAGATTTAAGGAAATGCTGAATAGCGCGG	720
Qy	721	CCAGATCCGGGAGGACAAATGAGAATTTCTTCCGCTGATCAAGTCCGAGCGAGACTGGCACTTTT	780
Dp	721	CCAGATCCGGGAGGACAAATGAGAATTTCTTCCGCTGATCAAGTCCGAGCGAGACTGGCACTTTT	780
Qy	781	GTTAATGGGTTTCCCGCAACGAGCGCTTCTTCTACTTTGAAATATATGAAACCGTGTAAAA	840
Dp	781	GTTAATGGGTTTCCCGCAACGAGCGCTTCTTCTACTTTGAAATATATGAAACCGTGTAAAA	840
Qy	841	GAAATCCGTTTACCCGACCGATGATCCGATCATCATGAACTGCGCGGACGAGCCGTGTC	900
Dp	841	GAAATCCGTTTACCCGACCGATGATCCGATCATCATGAACTGCGCGGACGAGCCGTGTC	900
Qy	901	GATCGGACACAGCAATTTGCAAAATGACGTAATGCGCTTATATTGAAAGCAAGTGAATCTGATC	960
Dp	901	GATCGGACACAGCAATTTGCAAAATGACGTAATGCGCTTATATTGAAAGCAAGTGAATCTGATC	960
Qy	961	TTAATTCGGTGAACCAACGAAACCGATCGTAAAGCTCTATGAAAGGCAAAATTTCTTGCA	1020
Dp	961	TTAATTCGGTGAACCAACGAAACCGATCGTAAAGCTCTATGAAAGGCAAAATTTCTTGCA	1020
Qy	1021	GACAAACTGCATGACCTAGAGTATATAGTCAACAGATGAATAATTATGGAATTTGTTAAAGAAA	1080
Dp	1021	GACAAACTGCATGACCTAGAGTATATAGTCAACAGATGAATAATTATGGAATTTGTTAAAGAAA	1080
Qy	1081	AGAAATGCAACAACGTCATATATATGCGTCGCGCAATATTCATATGTCGCGGACAGACTTTTA	1140
Dp	1081	AGAAATGCAACAACGTCATATATATGCGTCGCGCAATATTCATATGTCGCGGACAGACTTTTA	1140
Qy	1141	ATTGAAAAAATCCACGAAATACAAAGGTAAAGCAGCTCTGTAAAC	1182
Dp	1141	ATTGAAAAAATCCACGAAATACAAAGGTAAAGCAGCTCTGTAAAC	1182

```

1  Publication No. US20050255478A1
2  GENERAL INFORMATION:
3  APPLICANT: KIMBERLY, WILLIAM JOHN
4  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
5  FILE REFERENCE: PU3480US
6  CURRENT APPLICATION NUMBER: US/10/793,626
7  CURRENT FILING DATE: 2004-03-04
8  PRIOR APPLICATION NUMBER: 60/164,258
9  PRIOR FILING DATE: 1999-11-09
10 NUMBER OF SEQ ID NOS: 4472
11 SOFTWARE: PatentIn Ver. 2.1
12 SEQ ID NO: 3743
13 LENGTH: 3240
14 TYPE: DNA
15 ORGANISM: Artificial Sequence
16 FEATURE:
17 OTHER INFORMATION: Description of Artificial Sequence: synthetic
18 OTHER INFORMATION: nucleic acid sequence
19 US-10-793-626-3743

```

Query Match	28.6%; Score 338.6; DB 8; Length 3240;
-------------	--

Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;

5 GCTGTTACTATTATAGCCTGTGCTG--TCATACTGGTCATCGGAATATTAGAAAAC 61

Db 1673 GATTGTTACTCATTATAGCGTGTGTTGCGCTTATTCTTTGGCTAGGAATTAAAGAGAAGA 177

62 GACGACATCAGAAAAACATTGATGCCCTCCCTGTTGGGTGATATTAA CGCATCCGG 12

Db 1733 AGCGTCATGCAATCGACTTGAATAATTCATTACGTAATTACATAAATGGAATTCGTG 177

QY 122 GAAATCGACTGTGACAAGGCTGACCAACCGAATATTAATAGAGCCGGTTACACAGACTG 18

Db 1793 GGAATCTACGATTACTGGATGGCTTACAGTGTCTACGTGAGGACCAGTATAGAGTTG 18

182 TTGGAATAACACAGGACAGATGCAAGATTGATTACTGGACACACCGGAGGAAAGC 24

Db 1853 TTGGTAAGACCAACAGGACAGATGCACGAATGCTCTATTGGTTTACTGAAAGGAATATC 19

242 CGATTAAACGGAACCTCAGGGCCGAATATCGGAGAGCAAAAAGAAGTCATGAGAGAA 30

Db 1913 CAGTAATCAGAAACCACAAAGGCCCAACATTGGAGACACACGAGATATTTCGTAAGG 19

302 CAGTAGAAGAGGGGCTAACGGGATTTGTCAGTGAATGCATGGCTGTTAACCCAGATTATC 360

Db 1973 TGCTTAACAAAAGCAATGCCCTAGTTAATGAGTGATGGCTGTAATCCAGATTATC 20

362 AAATCATCTTTCAGGAGAACTTCTGCAGGCCAATATCGCGTCATTGTGAATGTTTAG 422

Db 2033 AAATTACCTTTCAAATGATTAGTAAAGCAATATCGGTGTAATTGTTAATGTGATGG 20

422 AAGACCATATGATGTCATGGGCCGACGCTTGATGAATTGCAGAAGCCTTACCGCTA 48

Db 2093 AAGACCATATGATGTTCTTAGGACCGACACTTAAGATGTAGCGCAAGCTTTACTGCA 21

482 CAATTCCTATAATGGCCATCTTGTCAATTAAGATAAGTAGAATATACCGAGTTCTTAAC 54

Db 2153 CAATTCATATAACGGGAATTAGTGTGTAATGAAGATACTATAGTTTCTTGCAA 22

542 AAAAAGCAAAGAACGAACACAAAGTCATCTTGTCTGATACTCAAATAATACAGATG 60

Db 2213 AGGAGCTAAAGCGTAATCAGAACTCATTGTTGTAGATAAAGACGTCATACCAGAAT 22

602 AGTATTACGTAATTTGATAACATGGTATTCCTGATAACGCTTCTCTGGCGCTGGGTG 66

Db 2273 CATATTACGAAGTTCGATTATTAGTATTCTGATAATGCTATTGTGTAGGA 23

662 TGGCTCAAGCACTCGGCATTGACGAGAACAAGCATTTAAGGGAATGCTGAATGCCCGC 72

Db 2333 TAGCGCAAGCAGTTGGTGTAGATGAAGAACTGCATTACACAGGTATGTTAAATGCACCAG 2334

722 CAGATCCGGGAGCAATGAGAATTCTCCGCTGATCAGTCCGAGCGAGCCTGGGCACTTG 78

RESULT 2
US-10-793-626-3743
; Sequence 3743, Application US/10793626

DB 2393 CGATCCAGGTCGTGTGAAATTAATTTCCATGCAATCGCAAAAATGATTTTG 2452
QY 782 TTAATGGTTGCCCAACGACGCTCTTCTACTTTGTAATATGAAACGTGTAAG 841
DB 2453 TTAATGCAATTCGCTGTAATGAAACGACGTACAAAAGCAATTTTAATTAAGTGAAT 2512
QY 842 AAATCGTTACCCGACCGATGATCCGATCATCATGAACTGCGCGACGCGTTCG 901
DB 2513 CATATAATTCATCATACGTAAGAAATATATCTCATTTGTGTCAGATGAGGTTG 2572
QY 902 ATCGACACAGCAATTCGCAATGAGTAATGCTTATATGAGCAAGTGAATGATCT 961
DB 2573 AATGACACCACTCTTTGTTGTAATCTTTTAGTGAAGTCATTAACATGTTTCATTT 2632
QY 962 TAATCGTGAACACAGAACCGAT 986
DB 2633 GTACAGAAAAAGTACACAAATGCT 2657

RESULT 3
US-10-793-626-2321

; Sequence 2321, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 2321
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2321

Query Match 20.8%; Score 246.2; DB 8; Length 861;
Best Local Similarity 60.0%; Pred. No. 2e-51;
Matches 410; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 304 GTAGAAAGAGGGCTAACCGGATGTCAGTGAATGATGCTGTTAAACCAATTTTCA 363
DB 4 GTTAAACAAAAGCAATGCTAGTAAAGTGAAGTGAATGCTGTTAAATCCAGATTTTCA 63
QY 364 ATATCTTTTCAGAGAGACTTTCGAGGCCAATATGCGCTCATTTGTAATGTTTAA 423
DB 64 ATTAACCTTTAAATGATTTAGTAAAGCAAAATATCGGTAAATTTGTAATGTAAGAA 123
QY 424 GACCATATGATGTCATGAGGCGCAGCTTGATGAATTCAGAGGTTTACCGTACA 483
DB 124 GACCAATATGATGTCATGAGGCGCAGCTTGATGAATTCAGAGGTTTACCGTACA 183
QY 484 ATTCCTTATATGAGCCATCTTGTTCATTAAGATAGTGAATATCCAGAGTTCTTAAACA 543
DB 184 ATTCCTTATATGAGCCATCTTGTTCATTAAGATAGTGAATATCCAGAGTTCTTAAACA 543
QY 544 AATGCAAAAGAGAAACAAAGTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 603
DB 244 GAAGCTAAAGAGTAAATTCAGAACTATGTTGTATGTAATGAAGCGTATACAGATCA 303
QY 604 TATTAGTAAATTTGATACATGATGATCCCTGATTAACGCTTCTGCGCGCTGGTGTG 663
DB 304 TATTAGGAAAGTTCGATTTATTTAGTATTTCTGATATATGATGATTTGTTAGGAATA 363
QY 664 GCTTACGACTCGGCTATGAGAAAGACGATTTAAGGAATGCTGAATGCGCCGCCA 723

DB 364 GCGAAGCAGTTGGTGTATGATGAAGAACTGCAATTAACAGGATGTTAATGACCAACC 423
QY 724 GATCCGGAGCAATGAAATTTCTCCGTGATGATCGAGCGAGCTGGGCACTTGTG 783
DB 424 GATCCAGGTGCTGTAGATTAATTAATTTCCATGCAATTCGACAAAATATGATTTGT 483
QY 784 AATGGTTTGGCCGAAACGACGCTTCTTCTACTTTGTAATATATGAAACGTGTAAGAA 843
DB 484 AATGATTCGCTGCTAATGAACCGAGCTTACAAAAGCAATTTTAATTAAGTGAATCA 543
QY 844 ATCGTTACCCGACCGATGATCCGATCATCATGAACTGCGCGACGAGCGTTCAT 903
DB 544 TATAATTCATATCGATTAAGAAATATATCTCATTTGTGTCAGATGAGGTTAT 603
QY 904 CGGACACAGCAATTCGCAATGAGTATGCTTATATGAGCAAGTGAATGATGCTTAA 963
DB 604 AGAACACCACTCTTTGTGATTAACCTTTTAGTGAAGTCATTAACATGTTCTCATTT 663
QY 964 ATCGGTGAACACAGAACCGAT 986
DB 664 ACAGAAAAAGTACACAAATGCT 686

RESULT 4
US-10-793-626-4090

; Sequence 4090, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 4090
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4090

Query Match 3.4%; Score 39.6; DB 8; Length 4642;
Best Local Similarity 43.6%; Pred. No. 9.6;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 237 AAAGCCGATTTAAAGGAACTCAGGCGCCGAATATCGAGAGCAAAAAGAACTCATG 296
DB 2052 AATGCAATTTAAAGTGAATTAATGTTTGTGTAATTTGCTGTTTGCATTCGAAG 2111
QY 297 AGAAAGATGAAGAGGGGCTTAAGCGATGTCAGTGAATGATGATGCTGTTAACCA 356
DB 2112 AATTCAGATGTAAGAGGCTTGAAGTATGATGATTAACGACTTAAACATGATCAT 2171
QY 357 TTAACAATCATCTTTCAGAAAGAACTTTCGAGGCCAATATCGGCTCATGTTGTAAT 416
DB 2172 GTTAGCTCATTTTAAATATGATGATGATGATGATGATGATGATGATGATGATG 2231
QY 417 TTGAAGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476
DB 2232 TATCGAAGGTGATTCGCTGTGAAACGTTAAAGAAATTAATCATTTGATGAACCA 2291
QY 477 CGCTAATTTCTTAAATAGCGCATCTTGCATTAAGATGATGATGATGATGATGATG 536
DB 2292 TGGTAAATTAACATGAGGCGCATTTAGATATCAAGTATTTAGTATGATGATGAT 2351
QY 537 TAAACAAAAGCAAGAAAGCAACAAAGTCAATCATGCTGATTAACCAAAAATTA 596
DB 2352 TACTGATTAAGAAAGCAACAGCTCATGATGATGATGATGATGATGATGATGATG 2411

OY 598 GATGAGTATTACGTAATTTGATACATGCTATTCCTGATAAGCT 645
 Db 239 GGTAAGGATTGCTTARATTTTATCTTGAGGCCAAGATTAGTCT 286

RESULT 15

US-09-925-065A-319482/c
 ; Sequence 319482, Application US/09925065A
 ; Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
 TITLE OF INVENTION: Identification and Mapping of Single
 ; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
 ; FILE REFERENCE: 108827.135
 ; CURRENT APPLICATION NUMBER: US/09/925,065A
 ; CURRENT FILING DATE: 2001-08-08
 ; PRIOR APPLICATION NUMBER: US 60/243,096
 ; PRIOR FILING DATE: 2000-10-24
 ; PRIOR APPLICATION NUMBER: US 60/252,147
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/250,092
 ; PRIOR FILING DATE: 2000-11-30
 ; PRIOR APPLICATION NUMBER: US 60/261,766
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/289,846
 ; PRIOR FILING DATE: 2001-05-09
 ; NUMBER OF SEQ ID NOS: 957086
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 319482
 ; LENGTH: 552
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-065A-319482

Query Match 3.1%; Score 37.2; DB 6; Length 552;

Best Local Similarity 48.6%; Pred. No. 17; Mismatches 108; Indels 0; Gaps 0;

Matches 102; Conservative 0;

OY 408 TGTGATGTTTGAAGACCATATGATGATGCGGCGACGCTTGATGAATTCAG 467
 Db 452 TATGACCTAATACATTTTTCAGATATACAAAGCAATTTACATGATGATGAC 393
 OY 468 AGCGTTTACCGCTACATTCCTTATTAATGCGCATCTGTCAATTACAGATATGATATAC 527
 Db 392 AGATAAATTAAACTTAGCTGGAATGTATGTTTATACAGACACTGAAAAAT 333
 OY 528 CGAGTCTTTAAACAAAAGCAAAAGACGAACAAAGATCATCTGCTGATAC 587
 Db 332 GAAGTAGACAAAATGTAAACATTAATGACAAATTTGAAACATTAACAAAGCTTACACTG 273
 OY 588 AAAAATTACAGATGATTTTACGTAATTT 617
 Db 272 ATTAAATTAAGGTTATTTTAAATGT 243

Search completed: February 27, 2006, 11:38:27
 Job time : 713.304 secs

THIS PAGE BLANK (USPTO)


```
Db 305 TGGTTAAACAAAAGCAATGCGCCAGTGTATGAGTGTATGCGTGTAAATCAAGTTATC 364
Qy 362 AAATCATCTTTTCAGAGAACTTCTGCAGGCCAATATGCGCTCATTTGTGAATGTTTAA 421
Db 365 AAATTCACCTTTCAAAATATTTAGTAAAGCAAAATGCGTGTATTTGTATGTGAAGG 424
Qy 422 AAGACCATATGAGTGTATGAGGCGCGAGCTTGTATGAATTTGCAAGACGTTTACCGGTA 481
Db 425 AAGACCATATGAGTGTATGAGGCGCGAGCTTAAAGATGAGCGACGCTTTTACTGCA 484
Qy 482 CAATTCCTTAATATGCGCCATCTTGTTCATTAAGATGATATATACCGAGTCTTTAAAC 541
Db 485 CAATTCCTTAATATGCGGAAATTTAGTGTATGAAGAAATATACATATCTAGTTTCTTTG 544
Qy 542 AAAAAAGCAAAAGCAAAACAAAGATCATCTGCTGATTAACCTCAAAAATTACAGATG 601
Db 545 AAGAAAGCTAAAGAGCTAATTCAGAACTCATTTGTTGTATGAAGACGTCATACCAAGAT 604
Qy 602 AGTATTTAGTAAATTTGAATATCATGATATTCCTGATTAAGCGCTTCTGCGCGTGGTG 661
Db 605 CATATTTACGGAAGTTTCATTTATTTAGTATTTCTGATTAATGATCTATTTGTTAGAA 664
Qy 662 TGGCTCAAGCACTCGGCAATTCAGAAAGCAAGCAATTTAAGGAAATGCTGAATGCGCCG 721
Db 665 TAGCGCAAGCAAGTTGCTGATGAAGAAAGCAAGCAATTCAGGATGTTAAATGACACAG 724
Qy 722 CAGATCCGAGCAATGAGAAATTTCTCCGCTGATATGTCGAGCGACGCTGGCACTTTG 781
Db 725 CCGATCCAGGCTGCTTGAATTAATTTCCATGCAAAATGCAAAAAATGATTTG 784
Qy 782 TTAATGGGTTTGGCGCAACGACGCTTCTCTACTTTGAATATATGAAGCGTAAAG 841
Db 785 TTAATGCAATTCGCTCTATATGAAACCGCAGCTTCAAAAAGCAATTTAAATGAAGTGA 844
Qy 842 AAATCGTTAACCCGACCGATGATCCGATCATCATGATGAATCCGCGCAGACCGTGTG 901
Db 845 CATATTAATTCATACGATGAAGAAATATATCATTTCTCAATGTGCTTCAAGTAGGTTG 904
Qy 902 ATCGACACAGCAATTCGCAATGACGATTCCTTATATTTGAAGCAAGTGAATGATCT 961
Db 905 ATAGAACAACAACCTCTTGTGTGTAATCTTTTATGAGTAAGTCAATGATGTTCTCATTT 964
Qy 962 TAATCGTGAACACAGAACCGAT 986
Db 965 GTACAGAAAAAGTACACAAATGCT 989

RESULT 2
US-09-710-279-3743
; Sequence 3743; Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3743
; LENGTH: 3240
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3743
Query Match 28.6%; Score 338.6; DB 3; Length 3240;
Best Local Similarity 59.7%; Pred. No. 1.7e-86;
```

```
Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;
Qy 5 GCTGGTACTCATATATGCGCTGCTG---TCATACGTGTCATCGAATATTTAGAAAC 61
Db 1673 GATGTACTCATATATAGCGTGTGTTCGCTTATTTCTTGGCTAGAAATTAAGAGAA 1732
Qy 62 GACGACATCAAAAAACATTGATGCCCTCCCTGTTCCGGTGAATATTTACGGCATCCGCG 121
Db 1733 AGCGTCATGCAAAATCGACTGAAAAAATTCATATAGTATTAACATAAATGGAATTCGTG 1792
Qy 122 GAAATAGACTGTGACAAAGCTGCAACCGGAATATTTAATGAAGCCGTTTACAGACTG 181
Db 1793 GGAATCTACGATTAATCGATGCTTACAGTGTCTACGCTGAGGACAGATATAGAGTTG 1852
Qy 182 TTGAAAAAACAAGCAAGATGCAAGATGATTTACTGAGGACACACCGAGGAAAGC 241
Db 1853 TTGTTAAGACAAACGAAACGATGCAAGATGCTCTTATTTGGTTTACTGAAAAAGAAATTC 1912
Qy 242 CGATTTAAACGAAACCTCAGGCGCCGAATATCGAGAGCAAAAAAGAGTCAATGAGAA 301
Db 1913 CAGTAAATCAGAAACCAACAGGCGCCCAACATTTGAGAAACAAGATATTTATTGTAAGG 1972
Qy 302 CAGTAAAGAGGCGCTAACGCGATTTGTCATGATGATGCTGTTTAAACCAATATTC 361
Db 1973 TGGTTAAACAAAAAGCAAAATGCCCTAGTATATGATGTATGCGTGTAAATCCAGATTATC 2032
Qy 362 AAATCATCTTTACAGAAAGAACTTCGAGGCGCAATATCGGCGTATGTAATGTTTAA 421
Db 2033 AAATTTACCTTCAAAAGATTTAGTAAAGCAAAATCGGTTAATTTGTAATGTATG 2092
Qy 422 AAGACATATGATGTCATGCGGCGCGACGCTTGATGAATTCGAAAGCGTTTACGCTA 481
Db 2093 AAGACATATGATGTCATGAGACGACATTAAGATGAGCGCAAGCTTTTACTGCA 2152
Qy 482 CAATTCCTTAATATGCGCATCTTTCATTTACAGATAGTAATATCCGAGTCTTTTAAAC 541
Db 2153 CAATTCATATTAAGGAAATTTAGTGTATGAAGATTAATACTAATGATTTCTTTC 2212
Qy 542 AAAAAAGCAAAAGCAAAACCAAAAGTCATCTTGTGATTAATCAAAAATTACAGATG 601
Db 2213 AAGAAAGCTAAAAACGTAATTCAGAACTCATTTGTTGATGAATTAAGACGTCATACCAAGAT 2272
Qy 602 AGTATTTACGTAAATTTGTAATACATGATATTCCTGTAACGCTTCTTGGCGCTGGTG 661
Db 2273 CATATTTACGGAAGTTGATTTATTTAGTATTTCTGTAATATGATGATGTTAGGAA 2332
Qy 662 TGGCTCAAGCACTCGGCAATTCAGAAAGCAAGCAATTTAAGGAAATGCTGAATGCGCGC 721
Db 2333 TAGCGCAAGCAAGTTGCTGATGATGAAGAACTGCATTAACAAGATGTATTAATCACACAG 2392
Qy 722 CAGATCCGAGCAATGAGAAATTTCTCCGCTGATCAATCCGAGGAGCGCGGCACTTTG 781
Db 2393 CCGATCCAGGCTGCTTGAATTAATTTTCATGCAAAATCGCAAAAAATGTATTTG 2452
Qy 782 TTAATGGGTTTGGCGCAACGACGCTTCTTACTTTGAATATATGAAACGTTAAAG 841
Db 2453 TTAATGCAATTCGCTGCTAATTAACGCACTTCAAAAAGGATTTTAAATTAATGGAAT 2512
Qy 842 AAATCGTTAACCCGACCGATGATCCGATCATCATGAAGTTCGCGCAGACCGTGTG 901
Db 2513 CATATTAATTCATACGATGAAGAAATTAATCATTTCTCAATTTGCTTCAAGATGAGGTTG 2572
Qy 902 ATCGACACAGCAATTCGCAATGACGATTTGCTTATATTTGAAGCAAGATGATGATCT 961
Db 905 ATAGAACAACAACCTCTTGTGTATACCTTTTATGAGTGAAGTCAATGATGATGTTCTCATTT 2632
Qy 962 TAATCGTGAACACAGAACCGAT 986
Db 2633 GTACAGAAAAAGTACACAAATGCT 2657

RESULT 3
US-09-710-279-2321
```


Oy 371 TTCAGAGAACTCTCGAGCGCAATATCGGCTCATGTAATGTTTAGAGCCATA 430
Db 1155 RR 1096
Oy 431 TGATGTCATGGGCGGCGGCTTGTAAGAAATTGCA 465
Db 1095 RR 1061

RESULT 5
US-09-248-796A-2670
Sequence 2670, Application US/09248796A

GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT FILING DATE: US/09/248,796A
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 2670
LENGTH: 2352
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-2670

Query Match 3.8%; Score 45.4; DB 3; Length 2352;
Best Local Similarity 51.8%; Pred. No. 0.015; Mismatches 96; Indels 0; Gaps 0;

Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
Oy 500 ATCTGTGATTAACAGATAGTAATATACGAGTCTTTAAACAAAGCAAGAAAGCA 559
Db 587 ATATTGTCTAGCAATATATATGTAACATATTAGCTTAACACAGCTAAAGCA 646
Oy 560 ACACAAAGTATCTGCTGATTAACGAAATTAAGATGATTAATTAATTTTG 619
Db 647 ATCAAAATCAATTAAGTAAAGCAATGTTCAACTAAGTGGTCTTCCATG 706
Oy 630 AATACATGATTCCTGATTAACGCTTCTCTGCGCTGGTGGCTCAAGACCTG 679
Db 707 AATCTATTCAGATATCTAAACAAATTAATCAACAATTAATTAATGCTG 766
Oy 680 TTGACGAGAAACAGCATT 698
Db 767 GTACTGTAAACATCATT 785

RESULT 6
US-09-806-708B-22/c
Sequence 22, Application US/09806708B
Patent No. 6784342
GENERAL INFORMATION:
APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741
CURRENT FILING DATE: US/09/806,708B
PRIOR APPLICATION NUMBER: US 60/147,133
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.0
SEQ ID NO 22
LENGTH: 1141
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: promoter

LOCATION: (1)..(1141)
OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match 3.4%; Score 40.6; DB 3; Length 1141;
Best Local Similarity 9.6%; Pred. No. 0.26; Mismatches 266; Indels 2; Gaps 1;

Matches 54; Conservative 241; Mismatches 266; Indels 2; Gaps 1;
Oy 13 CTCATATAGCTGTGCTGCTCATCTGATCGAATATTAAGAAACAGCAGCATCAG 72
Db 619 MTVAKKVKKMDTCTTVDVADSVWVWYAAWRCRDVTRNTYCKSVASHVYWNNA 560
Oy 73 AAAAAATTGATGCCCTCCCTGCTGGGTAATTTAAGCGCATCCGGAAAAATGACT 132
Db 559 MWYRRYARWWSWABRTTRNNMMWSGBVRRMAGTMMWNNNNNDTRYYMMKKMAR 500
Oy 133 GTGCAAGGCTGACACCGGATTTATATGAAGCCGGTTACAAGACTGTGAAAAACA 192
Db 499 BTTVYDSMCVAKSMWGNWRAKMMWAAANNDAADHWTYWGNNTMMRRRAKMMN 440
Oy 193 ACAGAAACAGATGCAAGATGATTAAGTGAACACCGGAGAAACCGATTAAACGG 252
Db 439 AWCRRATCCNNNNNRACVHKKMRTWKYMKACNNNNBRKMYRVAMWTSRDTN 380
Oy 253 AAACCTGAGGCGCGAATATCGAGACAAAGAACTCATGAGAAACAGTAAAGA 312
Db 379 TDMMWMTSDMBHHVTVDTYMRBMNNNNNNNNNNBCKTSMWMMMDHNTHTGNNTW 320
Oy 313 GGGGCTAACGGATTCAGTGAATGCAATGCGCTTTAAACCAATTTCATATCTTT 372
Db 319 SAYMAAMSMWAAASBVYVWCMWMTYMGKTNNNNNNKMYVETKTVAMCNNRY 260
Oy 373 CAGAAAGACTTCGACGCGCAATATGCGGCTATGATGATGTTTAAAGAACATATG 432
Db 259 DTAWMBRANKYCAVAVWYI--BMYGKHMBWRABHBSMMWVKCKNTVSWHY 202
Oy 433 GATGTCATGGGCGCGAATATCGAATGTAATGTAAGACGTTTACCGTACATTCCTAT 492
Db 201 HAMRYBKABVAGCANNMKRMAHHHCATNNNNMMWVYMHMMHKGAATNNKTAB 142
Oy 493 AATGCCATCTTGATTAACAGATAGTATATCCAGTCTTTAAACAAAGCAAAA 552
Db 141 RDDBAHVKTYWVRVYVWCMCMWNAKAKVTKMKIMWYTTDRVYVANNTVGMMWRW 82
Oy 553 GAAGAAACAAAGTCATCAT 575
Db 81 CMWYMSNNRMYRMRKTYWAW 59

RESULT 7
US-09-806-708B-22
Sequence 22, Application US/09806708B
Patent No. 6784342
GENERAL INFORMATION:
APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741
CURRENT FILING DATE: US/09/806,708B
PRIOR APPLICATION NUMBER: US 60/147,133
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.0
SEQ ID NO 22
LENGTH: 1141
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(1141)
OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match 3.4%; Score 40.4; DB 3; Length 1141;
Best Local Similarity 9.7%; Pred.No.0.3;
Matches 71; Conservative 285; Mismatches 373; Indels 2; Gaps 1;

```
406 ATGTGATGTTTGAAGACCATATGATGATGGGCGACGCTTGAATTAATTCGA 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
71 WYNNKSRWKKWKKYKKBKANTSBKTHARKKMKTAIBMTNNKKGKWRHRYRW 130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
466 GAAGCGTTTACCGCTACATTCCTTAATAAGCCATCTTGTCAATACAGATGTAAT 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
131 RMBDTVDHYYTANNNMTTTCMDKDKRTBMMWKKNNATGDDDTKYMMNNNGCB 190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 ACCGAGTCTTAAACAAAGCAAAAGCAAAAGTATCATCTGCTGATAC 585
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
191 TTTWVVRXYTIDMSBKNNYGBMWKWSYDVTYVWWWDKCKRYARWRTGRMKN 250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
566 TCAAAAATTACAGATGATTAATTAATTTGAATACATGATATCCCTGATTAACGT 645
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251 YVABTARRRYNNWYBAMAYRRKTMNNNNNNNNKAKKRYGMNRYAVNSTCTTWK 310
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
646 TCTCTGCGCTGGTGTGCTCAAGCACTGGCAATTGACGAAGAAACAGCATTTAAGGA 705
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
311 SKTTVTRTSQWANNCRAGDANKDHKKWKYSAAAGVYNNNNNNNNNTYKKAHBAWDMW 370
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
706 ATGCTGAATCGCCCGCCAGATCCGGAGCAATGAGAAATTCCTCCGTGATCAGTCCGAGC 765
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
371 HSAWKMHANAAHYSKKWTBYRKTMVNNNGTTWKKMMWAMWYKMDMBGTNNNNN 430
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
766 GAGCCGG--GGCACTTGTAAATGGTTGGCGCAAAAGCAAGCTTCTTACTTTGAATA 823
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
431 GRTTYGTNKKKKMYTKKANNCKRAMDHKTCTHNNNTTMMKMTYNNNCYKWSMTNG 490
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
824 TATGGAACGTTAAAGAAATCGTTTACCAGCCGATGATCGATCATCATGAACT 883
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 KHRBAAAYTYMMWRRYAHAANNMMWYMKACCTYKYBVCCKMNNYAAVYTTSSNN 550
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
884 GCCCGCAGACCGTGTGATCGGACACAGCAATTCGCAATGACGTAATTCCTTAATG 943
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
551 YTSRYRMKTNNNSWRSDTRSMGRANNYARABHYGYKNNTRWBSHTBHBRAGAHY 610
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
944 AAGCAAGTACATGATCTTAATCGGTGAACAACAAGCAAGCATGTAAG 1003
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
611 WMBMYBACHCKWAKYKAKYTAGAGSNNNNNNNNNNNNNNNNATCADDYTAASRWTA 670
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 AAGGCAAAATTCCTGAGCAAACTGATGACCTAGATGATAAGTACAGATGAATTA 1063
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
671 MANAKYTYKBAANAYTHANNWGCNNATDTRTMMWNNNNNNNAGTMMNNNNNAK 730
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1064 TCGAATTTGTAAGAAAGATGACACACCGTGTCAATATGCGCGCATATTCATG 1123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
731 ASAAKYVAAAAYKAKHWRMANKMARGHADAAABTDKRNNGAYTKYTTNNNNYTA 790
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1124 GTGCCGACAG 1134
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 791 GVVNTTAARDG 801
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 8
US-09-134-001C-2676
Sequence 2676, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 2676
LENGTH: 1047
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2676

Query Match 3.4%; Score 39.6; DB 3; Length 1047;
Best Local Similarity 43.6%; Pred.No.0.48;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```
237 AAAGCGATTAAACGGAACCTCAGGGCCGAATATCGGAGCAAAAGAGTCATGAG 296
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
36 AATGCAATTAAAGTACGAATTAATGTTTGTAAGTAATGTTGATTGACATTCAGAG 95
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 AGAAACATGAAAGAGGGCTTAACCGCATTTGTCAATGCAATGCGTGTAAACCCAGA 356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
96 AATTCAGATGAGAGGCTTTGAAGTAGTGACGTTTAACGACTTAACAGATGACGATAT 155
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
357 TTATCAATCATCTTTCAGAAAGACTTTCAGGCCCAATATCGGCGTCATTGTGAATGT 416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 GTTACGCTATTATTAAATAGACTAGTACGATGAGTGTTCATGAGAGTTGAAGT 215
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 TTTAGAGACCATATGATGATGATGAGGCGCAGCGCTGATGAATTCAGAGAGCTTAC 476
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
216 TATCGAAGTGGATTCCGTGTGAACGCTAAAGAAATTAATCATTCGATGAACGAGATGC 275
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
477 CGCTCAATTCCTTATATGGCCATCTTGCTATTACAGATGTAATGTAACGAGTCTT 536
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
276 TGGTAAATTTACATGAGGCGATTAGATATGACGATGATTAATGATGATGATGTTCTA 335
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
537 TAAACAAAGCAAAAGCAAGCAACAAAGTCAATTCATGCTGATTAATCAAAATTAAC 596
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 TACTGATTAAGAAAGCAAGCAAGTCAATGATGATGATGATGATGATTAATATCTC 395
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 AGATGATTAATTAATGTTTGAATACATGATGATGATGATGATGATGATGATGATG 642
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
396 AGCTCAGTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 441
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 9
US-09-710-279-4090
Sequence 4090, Application US/09710279
Patent No. 6703432
GENERAL INFORMATION:
APPLICANT: KIMBERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PUS480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4090
LENGTH: 4642
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4090

Query Match 3.4%; Score 39.6; DB 3; Length 4642;
Best Local Similarity 43.6%; Pred.No.0.96;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```
237 AAAGCGATTAAACGGAACCTCAGGGCCGAATATCGGAGCAAAAGAGTCATGAG 296
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2052 AATGCAATTAAAGTACGAATTAATGTTTGTAAGTAATGTTGATTGACATTCAGAG 2111
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
297 AGAAACATGAAAGAGGGCTTAACCGCATTTGTCAATGCAATGCGTGTAAACCCAGA 356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2112 AATTCAGATGATGAGAGGCTTGAAGTATGATGATGATGATGATGATGATGATGATG 2171
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 357 TTATCAATCATCTTTCAGAGAACTTCTGAGGCCAATATGCGCGTCATTGTAATG 416
DB 2172 GTTAGCTCATTTAATAATGATGACTATGCAAGGTCGTTCTGAGAAAGTTGAAGT 2231
QY 417 TTTAGAAACCATATGATGTCATGCGGCCGACGCTTGATGAAATTCGAGAACGTTTAC 476
DB 2232 TATGGAAGGTGATTCGGTGTGAACGGTAAGAAATTAATCATTCGATGAACGATGC 2291
QY 477 CGTACATATTCCTTATATAGCGCATCTTGTATTACAGATGTAATATACCGAGTTCTT 536
DB 2292 TGGTAAATTCACATGCGGCGATTTAGATATGACGTATATTAAGATGTAATCTGTTCTA 2351
QY 537 TAAACAAAAGCAAGAAAGCAACAAACAAAGTCATCTGCTGTAATCTCAAAAATTAC 596
DB 2352 TACGATTAAGAAAGAAAGCAAGCTCACATGATGACAGTCTTAATAAAGATTAACTC 2411
QY 597 AGATGATATTATTAAGTATTTTGAATACATGATATTCCTCGATAC 642
DB 2412 AGCTCCAGCTAAAGTATGTAATAAACAATGTAATTCAACTAAC 2457

RESULT 10
5231168-1
; Patent No. 5231168
; APPLICANT: DIEZIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO:1:
; LENGTH: 3095
5231168-1

Query Match 3.3%; Score 39.2; DB 9; Length 3095;
Best Local Similarity 46.2%; Pred. No. 1;
Matches 168; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

QY 281 AAAAAAGATGATGAGAGAAACAGTAGAAAGAGGGCTAACGCCATTGTCAATGATCA 340
DB 878 AAGAAAAACAATAATCACAGATCAGTTGAAGAAATTCAGTAATGAGATGAAATTTG 937
QY 341 TGGCTGTTAACCCAGATATTAATCATCTTTCAGAGAACTTTCGAGGCCAATATCG 400
DB 938 AAGATGTTCAACAGTGAACATTTAGATCATMAAAAGTTGATCCAGAAATAGTAG 997
QY 401 GCGTCAATGTGAATGTTTGAAGACCATATGATGTCATGG---GGCCGACGCTTGATG 457
DB 998 AAGTTGAAGAAATTCCTTCAGAACTACATGAATAATGAAGTGGCTCATCCAGAAATTTGTTG 1057
QY 458 AAATTGCAAGAGCGTTTACCGCTACAAATTCCTTATATATGGCCATTTGTCTATACAGATA 517
DB 1058 AAATTGAGGAAGTTTCTCTGAAACCAATCAAAATTAAGAAATTTCAATTAATGAAG 1117
QY 518 GTGATATATACCGAGTTCTTTAAACAAAGCAAAAGCAAGCAACAAAGTCATCATG 577
DB 1118 ATGATTAAGATGACATATATTCAGCATGAATAGTGAAGATGAAGAAATTAATCTCCAGAG 1177
QY 578 CTGATATACCAAAATTAACAGATGATATTTAGTATTTGTAATACATGATATTCCTG 637
DB 1178 ATGATTAAGAAATGAAGAAATGTAACATGAATATGTAAGTTGAAGAAATTTCTACAGAGAG 1237
QY 638 ATAA 641
DB 1238 ATAA 1241

RESULT 11
US-09-710-279-3209
; Sequence 3209, Application US/09710279
; Patent No. 6703492
```

```
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3209
; LENGTH: 1011
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3209
```

```
Query Match 3.3%; Score 38.6; DB 3; Length 1011;
Best Local Similarity 43.5%; Pred. No. 0.92;
Matches 176; Conservative 0; Mismatches 229; Indels 0; Gaps 0;
```

```
QY 238 AAGCCGATTAACGGAACCTCAGGGCCGATATTCGAGAGCAAAAAGATCATGAGA 297
DB 1 ATGGCAATTAAGATGAATTAATGTTTGTGAATGTTGTTAGCATTCAGAGAGA 60
QY 298 GAAACATGAAAGAGGGGCTTAACGCATTTGTCAGTAATGATGCTGTTAACCCAGAT 357
DB 61 ATTCAATATGTAAGAGTCTTGAAGTAGTTCAGTTAAACACTTAACAGATGACATATG 120
QY 358 TATCAATATCATCTTTCAGAAAGAACTTCGACGCCAATATTCGGCGTCATTGTAATGTT 417
DB 121 TTAGCTCATTTAATAATACATGATCATGATCAAGTGTCTTCACTGGAAGATTTGAAGTT 180
QY 418 TTAGAACCATATGATGATGTCATGCGGCCAGCGCTGATATAATTGCAGAGCGTTTACC 477
DB 181 ATCGAAGTGGATTCGCTGTGAACGTAAGAAATTAATCATTCGATGTAACCGATGCT 240
QY 478 GCTACATTCCTTATATATGCGCATCTTGTCAATACATATGTAATATACCGATGCTTT 537
DB 241 GTTAATTTACATGCGGCGATTTAGATATGACGTATATTAAGATGATCTGTTCTAT 300
QY 538 AACAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 597
DB 301 ACTGATTAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
QY 598 GATGATATTTACGTAATTTTGAATATACATGATATTCCTGATTAAC 642
DB 361 GCTCAGCTAAAGGTGATGTAATAAACAATGTAATTCAACTAAC 405
```

```
RESULT 12
US-09-134-000C-2669
; Sequence 2669, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2669
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2669
```

Query Match 3.2%; Score 38.4; DB 3; Length 1368;
Best Local Similarity 48.2%; Pred. No. 1.2; Indels 116; Gaps 0;
Matches 108; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

391 GCCAATATCCGCGCATTTGTAATGTTTGAAGACCATATGATGATGAGGCGCAG 450
736 GCGCTGGAAGAGTCCCTTCGGAATTCAGAGAGAGATGATGATGATGATGAT 795
451 CTGTATGAATTCGAGAGCGTTTACCGTACATTCCTTATATGCGCATTTGTCAT 510
796 AATCTGTGAAGGAGAGAGATTTAGTCAATTTTGGGAGAGAGATATACGTTAAG 855
511 ACAGATAGTGAATATACGAGTCTTTTAAACAAAGAGAGAGAGAGAGAGAGAG 570
856 ACGTATATGATATTCGAGATTCGAGATTCGAGATTCGAGATTCGAGATTCGAG 915

571 ATCATTTGCTGATACCTCAAAATTTACAGATGATGATTTTACGTA 614
916 AATGCTCCTTAAAGCTGAGAAATTCGAGAAATTTTACAGAA 959

RESULT 13
US-09-134-001C-2329
Sequence 2329 Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2329
LENGTH: 1158
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2329

Query Match 3.2%; Score 38.2; DB 3; Length 1158;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

452 TTGATGAATTTGAGAGCGTTTACCGCTACATTCCTATATATGCGCATTTGTATTA 511
224 TTGTTGAATTTACGAAACAGATTCAGTTCTCTGTCCTTTCTTCTGTTGATATTA 283
512 CAGATGATGAATATACCGAGTTCTTTAAACAAAGAGAGAGAGAGAGAGAGAG 571
284 TTAATAAATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 343
572 TCATTCGCTATATCTCAAAATTTACAGATGATTTTACGTAATTTTGAATCATGAT 631
344 TCACATCAGTATTCAGATTTTAACTTAAGTGCTTAAGTCTGATCATATACCATTA 403

632 TCCCTGA 638
404 TACCTGA 410

RESULT 14
US-08-323-170B-1
Sequence 1 Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pf230

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

Query Match 3.2%; Score 38.2; DB 2; Length 9636;
Best Local Similarity 55.7%; Pred. No. 3.4;
Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

921 AATGACGATTCCTTATATGAGCAAGTGAATCTTAATCGGTGAACAGACA 980
7309 AATTTCTTTTAAATATGAGAAATATGACACTACTATTAGATGAGAGAGACA 7368
981 ACCGATCGTAAAGCCTATGAGAGCAAAATTCCTGAGACAACTGCATGACTTGA 1040
7369 AAAAAAATTAACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7428

1041 GTATAGTCA 1051
7429 TGATGATTA 7439

RESULT 15
US-08-954-441-1
Sequence 1 Application US/08954441
Patent No. 6116000
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pf230
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/954,441
:   FILING DATE: 20-OCT-1997
:   CLASSIFICATION: 424
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: US 08/323,170
:     FILING DATE: 13-OCT-1994
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: US 08/010,409
:     FILING DATE: 29-JAN-1993
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Einhorn, Gregory P.
:     REGISTRATION NUMBER: 38,440
:     REFERENCE/DOCKET NUMBER: 015280-113110US
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE: (415) 576-0200
:     TELEFAX: (415) 576-0300
:   INFORMATION FOR SEQ ID NO: 1:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 9636 base pairs
:       TYPE: nucleic acid
:       STRANDEDNESS: single
:       TOPOLOGY: linear
:   FEATURE:
:     MOLECULE TYPE: DNA (genomic)
:     NAME/KEY: CDS
:     LOCATION: 149..9556
:   US-08-954-441-1
:
: Query Match      3.2% Score 38.2; DB 3; Length 9636;
: Best Local Similarity 55.7%; Pred.No.3.4;
: Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
:
: QY 921 AAATGACGTATTGCTTATTTATGAAAGCAAGTGAAGTGAATCGTGAACACACAGA 980
:   |||||
: DB 7309 AAATTTCTTTTACTTATATAGAAAATATATGCACTACTATTAGATGAGAAAGACA 7368
:   |||||
: QY 981 ACCGATCGTAAAGCCTATGAAGAGCAAAATTCCTGCAGCAAACTGCATGACCTAGA 1040
:   |||||
: DB 7369 AAAAATAATTAACACACACAGAGAGACACACACAAATATTTAAAGATCAAGA 7428
:   |||||
: QY 1041 GTATAGTCAA 1051
:   |||||
: DB 7429 TGATGATTAA 7439
:   |||||
```

Search completed: February 27, 2006, 06:46:09
Job time : 287.342 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 2573.57 Seconds
(without alignment)
9873.061 Million cell updates/sec

Title: US-10-789-164-2.

Perfect score: 447
Sequence: 1 atgttcgacgacgattata.....tcttcttactactact 447

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_da.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sta.*
11: gb_sy.*
12: gb_un.*
13: gb_vl.*
14: gb_hcg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	445.4	99.6	2775	1 DQ086153	DQ086153 Bacillus
2	443.8	99.3	4315	1 AB039950	AB039950 Bacillus
3	442.2	98.9	2989	1 AB016245	AB016245 Bacillus
4	442.2	98.9	8189	1 BSZ92954	BSZ92954 B. subtilis
5	442.2	98.9	200690	1 BSUB0019	BSUB0019 Bacillus su
6	440.6	98.6	450	6 ES0426	ES0426 Process for
7	440.6	98.6	3045	6 ES0426	ES0426 Process for
8	435.8	97.5	3738	1 AB046355	AB046355 Bacillus
9	290.8	65.1	110000	1 AB017333	AB017333 36
10	280.8	65.1	110000	1 CP000002	CP000002_36
11	289.2	64.7	450	6 AX434510	AX434510 Sequence
12	237.4	53.1	3244	1 BACCAPAB	BACCAPAB
13	237.4	53.1	94829	1 AB011191	AB011191 Bacillus
14	237.4	53.1	94830	1 AB011735	AB011735 Bacillus
15	237.4	53.1	96231	1 AF188935	AF188935 Bacillus
16	223.8	50.1	930	1 AB125962	AB125962 Bacillus
17	199.4	44.6	1336	1 BFU60883	BFU60883 Bacillus fi
18	199	44.5	110000	1 AP006716	AP006716_05

c 19	191	42.7	300275	1 AE016751	AE016751 Staphyloc
c 20	189.4	42.4	453	6 AR484743	AR484743 Sequence
c 21	189.4	42.4	453	6 AX143601	AX143601 Sequence
c 22	189.4	42.4	110000	1 CP000029	CP000029_21
c 23	182	40.7	3240	1 AF269703	AF269703 Staphyloc
c 24	182	40.7	3240	1 AR485657	AR485657 Sequence
c 25	182	40.7	3240	6 AX145021	AX145021 Sequence
c 26	180.4	40.4	110000	1 AP008934	AP008934 03
c 27	118.8	26.6	110000	1 BA000028	BA000028_02
c 28	53.8	12.0	110000	1 AE009951	AE009951_15
c 29	45.6	10.2	3029	1 AF270157	AF270157 Staphyloc
c 30	45.6	10.2	3029	6 AR486111	AR486111 Sequence
c 31	45.6	10.2	3029	6 AX145475	AX145475 Sequence
c 32	45.6	10.2	110000	1 AJ749949	AJ749949_08
c 33	43.2	9.7	158981	14 AC016242	AC016242
c 34	43.2	9.7	162151	8 AC012358	AC012358
c 35	41.2	9.2	125303	15 AC148304	AC148304
c 36	41.2	9.2	125303	15 AC147009	AC147009 Medicago
c 37	41.2	9.2	127018	15 AC135414	AC135414 Medicago
c 38	39.8	8.9	15319	15 AB023035	AB023035 Arabidops
c 39	39.6	8.9	221912	9 AL672003	AL672003 Mouse DNA
c 40	39.6	8.9	349980	6 AX344554	AX344554 Sequence
c 41	39.4	8.8	160286	15 AY780259	AY780259 Eucalyptu
c 42	39.4	8.8	175820	8 AL162378	AL162378 Human DNA
c 43	39.2	8.8	86901	9 AC105630	AC105630 Rattus no
c 44	39.2	8.8	200289	9 AC122295	AC122295 Mus muscu
c 45	39.2	8.8	232930	9 AC133761	AC133761 Rattus no

ALIGNMENTS

RESULT 1 DQ086153 2775 bp DNA linear BCT 27-JUN-2005
DQ086153 Bacillus subtilis strain ZJU-7 Pgab (pgab), Pgac (pgac), and Pgaa
(pgaa) genes, complete cds.

DEFINITION DQ086153.1 GI:68138277
Bacillus subtilis
Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE 1 (bases 1 to 2775)
Shi, F., Xu, Z., and Cen, P.
Efficient production of poly(glutamic acid) by a new strain
Bacillus subtilis ZJU-7

JOURNAL Unpublished
2 (bases 1 to 2775)
Shi, F., Xu, Z., and Cen, P.
Direct Submission
Submitted (04-JUN-2005) Institute of Bioengineering, Department of
Chemical Engineering and Bioengineering, Zhejiang University, No.38,
Zheda Road, Hangzhou, Zhejiang 310027, P.R. China

FEATURES
source
1..2775
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="ZJU-7"
/db_xref="taxon:1423"
1..1182
/gene="pgab"
1..1182
/gene="pgab"
/function="Biosynthesis of poly(glutamic acid)"
/codon_start=1
/product="Pgab"
/protein_id="AY85940.1"
/translation="MWLLIACAVILVIGLEKRRHQKIDALPVAVNINIGRKSTV
TRLTGLILEAGYKIVGKTGIDARMIYDTEPEKIKRKPQPNIGGEKVEKVE
RGNALVSECMVNPYQIIIFQELLQANIGVTAVNLIEDHMDVGPFLDEAFATAT

IPNGHLVITDSEYTERFEKOKAKERNTKVIILADNSKITDEYLRKEFWVPDNLSAL
GVAOLGIDEETAFKGMNLAPDPGAMLIPLISSEBGFVNGFANDASTLNIWK
RKEIYGPETDPIIIMNCRADRVDRTOOFANDVLPYIEASELLIGETTERIVAYEE
GKIPADKLDHLEYKSTDEIMELKKRMHNRIYGVGNHGAEPLEIKIHEYKVKQLY
S"

gene
CDS

1183 . 1632
/gene="pgac"
1183 . 1632
/gene="pgac"
/function="biosynthesis of poly(glutamic acid)"
/codon_start=1
/transl_table=11
/product="Pgac"
/protein_id="AAV85941.1"
/db_xref="GI:68138279"
/translation="MFGSDIYALIGVLSLIPAEKTIYVAGLVVPGYLVGNOP
VFILVLVSLTIVYIVKXGSKPMILYGRKRFAMMLITGVLYLKAIFPLIYFPEI
AEFRIGIIVPELIANTIOKQGLITTFGSILLSGATFALMFVYLLI"
1633 . 2775
/gene="pgaa"
1633 . 2775
/gene="pgaa"
/function="biosynthesis of poly(glutamic acid)"
/codon_start=1
/transl_table=11
/product="Pgaa"
/protein_id="AAV85942.1"
/db_xref="GI:68138280"
/translation="MKKELSFHEKLKLTROQKKTKNHVFIPIVFLMFAFMAG
KAETPKVTVSDVLSASFVGDIMMGRVETKEGKADSIFQVEPIFRASDVAGN
FENPVYOKNYKQADKEIHLQTNKESYKVLKDMNTVLSANNHAMDYGVQGMKDTLG
EFKQNLDIVGAGYSLSDAKKISYQVNGVTLATLTFVDSGKFAKNTPGVLLPA
DPEIFIMISEAKKHADIWVVOHMGGEYNDPNDROQLARASDAGADIIVGHHPH
VLBEIYVNGTVIIFYSLGNFVFDQGWTRBDSALVYHLKKNGTGREVTPIDTHEAT
PAPVKDSLKQKTIIRLETKDSNFAKVEDKLTVDHSDKLSK"

gene
CDS

1 ATGTTCCGATCGATTATATACATCGCACTAATTTTAAAGTGATCTACTGATTTAATTTT 60
1183 ATGTTCCGATCGATTATATACATCGCACTAATTTTAAAGTGATCTACTGATTTAATTTT 1242
61 GCGGAAAAACAGGAGATCGTCCGCGAGACTTGTGTACCCGGATATTAGACTTGTG 120
1243 GCGGAAAAACAGGAGATCGTCCGCGAGACTTGTGTACCCGGATATTAGACTTGTG 1302
121 TTTAATCAGCGGCTTTTATTTTCTGTTTGTAGTGAAGCTGCTCACTAATGTTATC 180
1303 TTTAATCAGCGGCTTTTATTTTCTGTTTGTAGTGAAGCTGCTCACTAATGTTATC 1362
181 GTGAATACGGTATTTCATAATTTGATTTTGTACGACGAGAAAATTCGCTCCATG 240
1363 GTGAATACGGTATTTCATAATTTGATTTTGTACGACGAGAAAATTCGCTCCATG 1422
241 CTGATTAACAGGAGATGCTCTAATAATCGCGTTTGTATTTCTATACCCGATTTGACATT 300
1423 CTGATTAACAGGAGATGCTCTAATAATCGCGTTTGTATTTCTATACCCGATTTGACATT 1482
301 GAAATCGGAGATTTTCAGGAAATCGGCAATCGAGCCAGGTTAATGCGAATACCAAT 360
1483 GAAATCGGAGATTTTCAGGAAATCGGCAATCGAGCCAGGTTAATGCGAATACCAAT 1542
361 CAGAAACAGGTTTAAACATTAGTTCGAGACAGCTGCTATTGAGCGGACGACCTTT 420
1543 CAGAAACAGGTTTAAACATTAGTTCGAGACAGCTGCTATTGAGCGGACGACCTTT 1602
421 GCTATCATGTTTGTATTACTTAAAT 447
1603 GCTATCATGTTTGTATTACTTAAAT 1629

ORIGIN

Query Match 99.6%; Score 445.4; DB 1; Length 2775;
Blast Local Similarity 99.8%; Pred. No. 98-105;
Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 2
AB039950
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES

source

gene
CDS

gene
CDS

gene
CDS

AB039950 4315 bp DNA linear BCT 12-SEP-2000
Bacillus subtilis capB, capC, capA, ywec genes, complete cds.
AB039950
AB039950.1 GI:10119860
CAPA; CapC; CapB.
Bacillus subtilis
Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 4315)
Tran, L.P. and Itoh, Y.
Nucleotide sequence of the capBCA operon in Bacillus subtilis
(natto)
2 (bases 1 to 4315)
Tran, L.P. and Itoh, Y.
Direct Submission
Submitted (08-MAR-2000) Yoshifumi Itoh, National Food Research
Institute, Applied Microbiology; Kannondai 2-1-2, Tsukuba, Ibaraki
305-8642, Japan (E-mail: yosifumi@nfri.affrc.go.jp,
Tel: +81-298-38-8075, Fax: +81-298-38-7936)
Location/Qualifiers
1 . 4315
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="Asehi-kawa"
/db_xref="taxon:1423"
941 . 2122
/gene="capB"
941 . 2122
/gene="capB"
/codon_start=1
/transl_table=11
/product="CapB"
/protein_id="BAB13484.1"
/db_xref="GI:10119861"
/translation="MMLIIACAVILVIGLEKRRHQNIDALPVRVNINGISGKSTV
TRLTGILTAGYKTVGKTGTGTGADARVYMTPEPKPIKRPQGNIDEGKEMRETV
RGANATVSECAVNPVQIIFQEBLOANIGVIVNVEDMDVNGPFLDEIARFTAT
IPVNGHLVITDSEYTERFEKOKAKERNTKVIILADNSKITDEYLRKEFWVPDNLSAL
GVAOLGIDEETAFKGMNLAPDPGAMLIPLISSEBGFVNGFANDASTLNIWK
RKEIYGPETDPIIIMNCRADRVDRTOOFANDVLPYIEASELLIGETTERIVAYEE
GKIPADKLDHLEYKSTDEIMELKKRMHNRIYGVGNHGAEPLEIKIHEYKVKQLY
S"
2137 . 2586
/gene="capC"
2137 . 2586
/gene="capC"
/codon_start=1
/transl_table=11
/product="CapC"
/protein_id="BAB13485.1"
/db_xref="GI:10119862"
/translation="MFGSDIYALIGVLSLIPAEKTIYVAGLVVPGYLVGNOP
VFILVLVSLTIVYIVKXGSKPMILYGRKRFAMMLITGVLYLKAIFPLIYFPEI
AEFRIGIIVPELIANTIOKQGLITTFGSILLSGATFALMFVYLLI"
2605 . 3747
/gene="capA"
2605 . 3747
/gene="capA"
/codon_start=1
/transl_table=11
/product="CapA"
/protein_id="BAB13486.1"
/db_xref="GI:10119863"
/translation="MKKELSFHEKLKLTROQKKTKNHVFIPIVFLMFAFMAG
KAETPKVTVSDVLSASFVGDIMMGRVETKEGKADSIFQVEPIFRASDVAGN
FENPVYOKNYKQADKEIHLQTNKESYKVLKDMNTVLSANNHAMDYGVQGMKDTLG
EFKQNLDIVGAGYSLSDAKKISYQVNGVTLATLTFVDSGKFAKNTPGVLLPA
DPEIFIMISEAKKHADIWVVOHMGGEYNDPNDROQLARASDAGADIIVGHHPH
VLBEIYVNGTVIIFYSLGNFVFDQGWTRBDSALVYHLKKNGTGREVTPIDTHEAT

gene
3762.. .3929
/gene="ywcC"
3762.. .3929
/gene="ywcC"
/codon_start=1
/transl_table=11
/protein_id="BAB13487.1"
/db_xref="GI:10119864"
/translation="MKFVAIWPFAVAIVFWMFSAFKENQDITDQEKQKIDMENKIKQOEPPVANK"

ORIGIN

Query Match 99.3%; Score 443.8; DB 1; Length 4315;
Best Local Similarity 99.6%; Pred. No. 2.2e-104; Mismatches 2; Indels 0; Gaps 0;
Matches 445; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTTGGATCGATTATATACATCGCACTAATTTAGTGTACTACTAGTTAATTTT 60
DB 2137 ATGTTGGATCGATTATATACATCGCACTAATTTAGTGTACTACTAGTTAATTTT 2196
QY 61 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGACCGGATATTTAGACTTGTG 120
DB 2197 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGACCGGATATTTAGACTTGTG 2256
QY 121 TTTAATCAGCCGGCTTTATTTACTGTTTCTAGTGAAGCTTCTCACTATATGTTATC 180
DB 2257 TTTAATCAGCCGGCTTTATTTACTGTTTCTAGTGAAGCTTCTCACTATATGTTATC 2316
QY 181 GTGAATACGGTTATCCAAATTTATGATTTTGTACGAGCGGAAAAATTCGTCGCATG 240
DB 2317 GTGAATACGGTTATCCAAATTTATGATTTTGTACGAGCGGAAAAATTCGTCGCATG 2376
QY 241 CTGATACAGGATCGTCTTAAATCGCGTTGATTTCTATACCGGATTTGACCATTT 300
DB 2377 CTGATACAGGATCGTCTTAAATCGCGTTGATTTCTATACCGGATTTGACCATTT 2436
QY 301 GAAATCGCAAAATTCGAGGAATCGGCATCATGTCGACAGTTAAATTCGCAATACCAT 360
DB 2437 GAAATCGCAAAATTCGAGGAATCGGCATCATGTCGACAGTTAAATTCGCAATACCAT 2496
QY 361 CAGAAACAGGTTAACCATTTAGTTCGAGACGCTGCTATTTGACGAGGACCTTT 420
DB 2497 CAGAAACAGGTTAACCATTTAGTTCGAGACGCTGCTATTTGACGAGGACCTTT 2556
QY 421 GCTATCATGTTTGTACTACTAATTT 447
DB 2557 GCTATCATGTTTGTACTACTAATTT 2583

RESULT 3
AB016245 2989 bp DNA linear BCT 14-OCT-1999
LOCUS Bacillus subtilis pgsA, pgsB and pgsC genes, complete cds.
DEFINITION
AB016245
AB016245.1 GI:6045071
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 (sites)
Ahiuchi, M., Soda, K. and Misono, H.
A poly-gamma-glutamate synthetic system of Bacillus subtilis IFO
3336: gene cloning and biochemical analysis of poly-gamma-glutamate
produced by Escherichia coli clone cells
Biochem. Biophys. Res. Commun. 263 (1), 6-12 (1999)
JOURNAL
PUBMED
10486244
2 (bases 1 to 2989)
AUTHORS
Ahiuchi, M.
TITLE
Direct Submission
JOURNAL
Submitted (14-JUN-1998) Makoto Ahiuchi, Kochi University, Research
Institute of Molecular Genetics; Otsu 200, Monobe, Nankoku, Kochi
783-8502, Japan (E-mail: ahiuchi@ring.kochi-u.ac.jp)

FEATURES
source
location/Qualifiers
1..2989
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="IFO 3336"
/sub_species="Bacillus natto"
/db_xref="taxon:1423"
1..1182
/gene="pgsB"
1..1182
/gene="pgsB"
/codon_start=1
/transl_table=11
/product="pgsB"
/protein_id="BA85263.1"
/db_xref="GI:6045072"

Tel: 0888-64-5215, Fax: 0888-64-5109)

gene
CDS

1197..1646
/gene="pgsC"
1197..1646
/gene="pgsC"
/codon_start=1
/transl_table=11
/product="pgsC"
/protein_id="BA85264.1"
/db_xref="GI:6045073"
/translation="MFGSDLYIALILGVLSLIPAEKTIIVPAGLVPEYGLVENOP
VETLVLLVSLTVIVIVKYGSKFMILYGRKFAAMLITGVILKIAFPLIVPEI
AEFRGIIIVPLIANTIOKQGLITPSTLLLSGATPAIMFVYLI"
1665..2807
/gene="pgsA"
1665..2807
/gene="pgsA"
/codon_start=1
/transl_table=11
/product="pgsA"
/protein_id="BA85265.1"
/db_xref="GI:6045074"
/translation="MKKELSFHEKLLKLTQOKKKTKNKHVFIAPIVFLMFAFMAG
KAETPKVTSDDVLSASFVDIMMGRVYEKTEQKADSIROYVEPIRADYVAGN
FENPVYOKNRYOKADKEIHLQTNKESVKLKMNFVLNSANNHAMDYVGQMKDTLG
EBAKQNDIVGAGYSLSDAKKISYQKNGVTIATLGFPTDVGSGKAPGKTVGLPA
DEEIPFPMISEAKKHADIVVDSHNGOEYRDNPNRQRLAAMSDAGADIIYGHHPH
VLEPIEVNGIVIFYSLSNFPDQWTRRDSALVOYHUKKGTGFETPTPIDHEAT
PAPVKDSLKOKTIIRELTKDSNFAKVEDGKLTFTDIDHSLKSK"

gene
CDS

ORIGIN

Query Match 98.9%; Score 442.2; DB 1; Length 2989;
Best Local Similarity 99.3%; Pred. No. 6e-104; Mismatches 3; Indels 0; Gaps 0;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTTGGATCGATTATATACATCGCACTAATTTAGTGTACTACTAGTTAATTTT 60
DB 1197 ATGTTGGATCGATTATATACATCGCACTAATTTAGTGTACTACTAGTTAATTTT 1256
QY 61 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGACCGGATATTTAGACTTGTG 120
DB 1257 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGACCGGATATTTAGACTTGTG 1316
QY 121 TTTAATCAGCCGGCTTTATTTACTGTTTCTAGTGAAGCTTCTCACTATATGTTATC 180
DB 1317 TTTAATCAGCCGGCTTTATTTACTGTTTCTAGTGAAGCTTCTCACTATATGTTATC 1376
QY 181 GTGAATACGGTTATCCAAATTTATGATTTTGTACGAGCGGAAAAATTCGTCGCATG 240

```
Db      1377 GTGAATAAGCGTTTATCCAAATTATGATTTGTACGAGCAGCAAAAATTCGCTGCCATG 1436
Oy      241 CTGATTAACAGGATGCTCTTAAATAATCGCGTTTGATTTTCTATACCCGATGTGACCATTT 300
Db      1437 CTGATTAACAGGATGCTCTTAAATAATCGCGTTTGATTTTCTATACCCGATGTGACCATTT 1496
Oy      301 GAAATCGCAGATTTGAGAGATCGGCATTCGCGACGAGTTTAAATGGCAATACCATTT 360
Db      1497 GAAATCGCAGATTTGAGAGATCGGCATTCGCGACGAGTTTAAATGGCAATACCATTT 1556
Oy      361 CAGAAACAGAGTTTAAACCATTAACGTTTCGAAAGCAGCGCTGTAATGAGCGGAGCGACCTTT 420
Db      1557 CAGAAACAGAGTTTAAACCATTAACGTTTCGAAAGCAGCGCTGTAATGAGCGGAGCGACCTTT 1616
Oy      421 GCTATCATGTTTGTCTTACTACTAATT 447
Db      1617 GCTATCATGTTTGTCTTACTACTAATT 1643

RESULT 4
LOCUS   BS292954 8189 bp DNA linear BCT 02-DEC-2004
DEFINITION B.subtilis ywe [A,B,C,D,E,F,G] and gerBC genes.
VERSION   292954
KEYWORDS  gerBC gene; ywsc gene; ywfa gene; ywcb gene; ywcc gene; ywcd gene;
SOURCE    ywE gene; ywcf gene; ywG gene.
ORGANISM  Bacillus subtilis
REFERENCE 1 (bases 1 to 8189)
AUTHORS   Prescan E., Mozer I., Boursier L., Cruz Ramos H.C., de la
          Fuente V., Hillo M.F., Lelong C., Schleich S., Sekowska A.,
          Song B.H., Villani G., Kunec F., Danchin A. and Glaeser P.
          The Bacillus subtilis genome from gerBC (311 degrees) to licR (334
          degrees)
          Microbiology (Reading, Engl.) 143 (Pt 10), 3313-3328 (1997)
          953393
          2 (bases 1 to 8189)
          Lejong C., Glaeser P., Prescan E. and Danchin A.
          Bacillus subtilis gerB downstream sequence
          Unpublished
          3 (bases 1 to 8189)
          Glaeser P.
          Direct Submission
          Submitted (13-MAR-1997) Philippe Glaeser, Regulation de l'Expression
          Génétique, Institut Pasteur, 28 Rue du Dr Roux, Paris, 75724,
          FRANCE

FEATURES
     source          1..8189
                        location/Qualifiers
                        /organism="Bacillus subtilis"
                        /mol_type="genomic DNA"
                        /strain="168"
                        /db_xref="taxon:1423"
                        1..1069
                        /gene="ywsc"
                        <1..1069
                        /gene="ywsc"
                        /note="product highly similar to Bacillus anthracis CapB
                        protein"
                        /codon_start=2
                        /transl_table=1
                        /protein_id="CAB07475.1"
                        /db_xref="GI:1894765"
                        /db_xref="GOA:Q798J2"
                        /transl_table=1
                        /translation="RGKSTVRLTGLIGGYKTVGKTGTGDAMIYMDTPEEKPIK
                        RKQSPNIGEOKEWRETVBERGANIVSECMANVPDYOITIOBEILQANIGIVYVLE
                        DHMDVMPGLDIAAFTATITPYNGHLVITDSEITPEFKQAKKENTVITITADNSKIT
                        DEYLKFEYVFPDNASIALGVQAQLGIDETAFKGMNAPPDGAAMILPILSPSE
                        GHFVNGFPAANDASSTINIMKRVKEIGYPTDPIIIMNCRADVDTQOFADVLPYIE
                        ASELILGETTEPIVKAEEGKIPADKLDLEYSKTDIMELIKKSMHNRVIYGVGNI
                        HGAAPFLIEKTHEYKVKQLVS"

     gene            join(1069..1075,1084..1533)
                        /gene="ywfa"
                        1069..1075
                        /gene="ywfa"
                        1084..1533
                        /gene="ywfa"
                        /note="product highly similar to Bacillus anthracis CapC
                        protein"
                        /codon_start=1
                        /transl_table=1
                        /protein_id="CAB07468.1"
                        /db_xref="GI:1894766"
                        /db_xref="GOA:P96737"
                        /db_xref="UniProt/TREMBL:P96737"
                        /translation="MEGSDLYIALILGVLSLFAEKTVLPAGLVVGYLGVENP
                        VETLVVLSLTVLVYKGLSKEMILYGRKRAAMITGIVKIAEFDPLVLPPEPI
                        AERFGIIVPGLIANTTOKOGLITTFGSTLLSGATFAIMFYILI"
                        1538..2694
                        /gene="ywcb"
                        1538..1544
                        /gene="ywcb"
                        1552..2694
                        /gene="ywcb"
                        /note="product highly similar to Bacillus anthracis CapA
                        protein"
                        /codon_start=1
                        /transl_table=1
                        /protein_id="CAB07469.1"
                        /db_xref="GI:1894767"
                        /db_xref="UniProt/TREMBL:P96738"
                        /translation="MKKLSFHSKLLKLTQOKKTKNHVFIPIVFLMFAFMWAG
                        KAETPKVYTSDDVLSAPVDIMMGHYVEKTEQKADSIQYVEPIFASDVAQN
                        FENPVYOKRYKADKEIHLQTKKESKVKLNAPFYIANNHAMQYQGMKDTLG
                        EPKQNDIDVAGAGSLSDAKKISYQKYNVTATLGTFTVSSKGFAPKNTGVLBA
                        DPETFIWSEAKHADVIVQSHMGSEYNDPNDRQRLARMSDAGADITVGHNRH
                        VLEIEVYNGTVIFYSLGNFVFDGWTRTDLSALVOYHLKNGRGREVPIDHEAT
                        PAPVKDLSKQITIRLTXDINSFAMVEDGKLTFDIDHSKJLKS"
                        2697..2914
                        /gene="ywcc"
                        2697..2702
                        /gene="ywcc"
                        2709..2876
                        /gene="ywcc"
                        /codon_start=1
                        /transl_table=1
                        /product="unknown"
                        /protein_id="CAB07470.1"
                        /db_xref="GI:1894768"
                        /db_xref="UniProt/Swiss-Prot:P96739"
                        /translation="MKFVKAIPFVAIVFMMSAFKFNDOULTDOKOKIDMEMNKI
                        QOQEEPPNANK"
                        2880..2914
                        /gene="ywcc"
                        3000..4258
                        /gene="ywcd"
                        3000..3007
                        /gene="ywcd"
                        3017..4258
                        /gene="ywcd"
                        /note="product similar to E. coli NUPC protein and to
                        listeria species p60-related proteins"
                        /codon_start=1
                        /transl_table=1
                        /protein_id="CAB07476.1"
                        /db_xref="GI:1894769"
                        /db_xref="GOA:P96740"
                        /db_xref="UniProt/Swiss-Prot:P96740"
                        /translation="MNTLANMKKELVAVILICFLVPTKAEIAEADTSSELIYSEAK
                        NLCGYKVGSETPKSGPDPSGLIOVFSKADILHPSVANOVYKIGTAVPENKPPD
                        ILFKKSGSTGYPTHDALYIGSQMHSQSKVITITVKKSSYNSGYTGARIKA
                        DPATADVPVQAEAKYIGVYVFGGSTPSGFDGSLVQVVFQDALGIYLPRAEQM
                        AVGKRVAPDNIKPDVVYFNTYKTGISHAGIYAGAGRTQASRSKRVTSIYSEDYW
                        KSKMTGIRRFNDNLITPRENDIVSEATLYGVEVPYKQGVLPETGFDAGVQVYVYORA
```


Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J., Sekowska,A., Seror,S.J., Serrot,P., Shin,B.S., Soldo,B., Sorokin,A., Tacconi,B., Takagi,T., Takahashi,H., Takemaru,K., Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpeira,P., Tognoni,A., Tosato,V., Uchilyama,S., Vandenbol,M., Vannier,F., Vassarotti,A., Viari,A., Wambuit,R., Wedler,E., Wedler,H., Weltzenegger,T., Winters,P., Wipac,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K., Yoshida,K., Yoshikawa,H.F., Zumbstein,E., Yoshikawa,H. and Danchin,A.

The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*

Nature 390 (6657), 249-256 (1997)

2 (bases 1 to 200690)

Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.

Direct Submission

Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur, Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

On Jul 7, 2003 this sequence version replaced gi:2636029.

This entry contains data from release R16.1 of the Subtilist database. Further data on gene annotation and detailed information about changes from previous releases can be found at <http://genolist.pasteur.fr/Subtilist/>.

location/Qualifiers

1. 200690

/organism="Bacillus subtilis subsp. subtilis str. 168"

/mol_type="genomic DNA"

/strain="168"

/db_xref="taxon:224308"

/complement(114. .2987)

/gene="uvrA"

/locus_tag="BSU35160"

/complement(114. .2987)

/gene="uvrA"

/locus_tag="BSU35160"

/function="excision of ultraviolet light-induced pyrimidine dimers in DNA"

/codon_start=1

/transl_table=11

/product="excinuclease ABC (subunit A)"

/protein_id="CAB15533.1"

/db_xref="GI:2636042"

/db_xref="GOA:O34863"

/db_xref="InterPro:IPR003439"

/db_xref="InterPro:IPR004593"

/db_xref="InterPro:IPR004602"

/db_xref="Subtilist:BG12697"

/db_xref="UniProt/Swiss-Prot:O34863"

/translation="WAMDRIEYKARAHNLKVIDITPRDDLVVVTGLSGSGKSLAFVTIYAEQGRRYVESLSAYARQPLGMDKPDVAIEGLSPAISIDQKTSRNPRTVGTVEIYDYLRLIYARVGPCHPEHGIETISQIEQWVDRILEPERTKLQVAPIVSGRKGAVYKLEQIRKOGYRVIRIDGEMAEISDIELEKNKHSIEPVLIRIVVEGAAKLSDSLFETRLIGEGRVMTDIVGEELMFSEHHAHPGCHGISGELEPRLIFSPSPGACPTCDLGMKLEVDADLVTPNODLSKENAVPMTISQVYRQLEAVTCHYGLDMYPIVDLPKQOLDKVLVLSGSDLLTYFRFENDPGQIRGELOFBSGLVKNIRRYKETSDFIREQMEQYMSQSCPTCKGKYLKKEALNDGRHIGITELSVADALAFEDLTISEKDMQIANTLIREIVERLSEFLDVGADVLISRAAGTLGGENQRIRLATQIGSRISGLYIIDEPSIGLHNDNRILISALKMNRDGNLIVVEHEDETMAADVLIDGPAGIHGGOVISAGTPEEWEDPNSLTGSYLSGKKFPLPERKKPGRGYIEIKGASENNLKKVNAKRPILGFTTAVTGVSGSGKSTIVNELIKRLAKLKKAKKPGSHKEIYGLHLDKVIDIOAPIGRTPRSNPATYTTTVPDIDIRVFAQTBEAKRGYKGRFFSNYKGCCEACRGDGIIRKIEHMFLDVYVPCVCHGKRYNRRETLVYKXSI SDVLMTYEDALSFEENIPIKIKRLQTLVGVGLGYITLQGPATTLTSGEQRVYKLSLHKRSTGRLLYLIDEPTTGHVDIARLLVLOELVNDGTVIENHIIIKTADYIVDLGEGGAGCGGTIVASGTPEELTEVESYTRYLKEVIRDRTRMKSLKAKETATS"

/complement(2995. .4980)

/gene="uvrB"

/locus_tag="BSU35170"

/complement(2995. .4980)

/gene="uvrB"

/locus_tag="BSU35170"

/function="excision of ultraviolet light-induced pyrimidine dimers in DNA"

/note="alternate gene name: dlnA, uvrA"

/codon_start=1

/transl_table=11

/product="excinuclease ABC (subunit B)"

/protein_id="CAB15534.1"

/db_xref="GI:2636043"

/db_xref="GOA:P37954"

/db_xref="InterPro:IPR001410"

/db_xref="InterPro:IPR001650"

/db_xref="InterPro:IPR001943"

/db_xref="InterPro:IPR004807"

/db_xref="InterPro:IPR006935"

/db_xref="InterPro:IPR009055"

/db_xref="Subtilist:BG10502"

/db_xref="UniProt/Swiss-Prot:P37954"

/translation="MKDRFELVSKYQPGQGPAAIEKLVKGIQEGKKQTLGATGTGKTFVSNLIKENVKPTLVIAHNKTLAGQLSEFEKPEPNNAVYFVSYYDYOPEAVYPQDTFIEKQASINDEIDKLHSAISLAFRRDVIIIASVCIGYSGPEEYMYVSLRTEIERREILRKVIDIOVANDIDFORGTVRGDDVIEIFPASRDEHCVVEFGDEIRIREVDALGELIGDHDHVAIFPASHFVTRAEKMEKAIQNIKELEEQKNVHENGLEAQRLEQRTVDLEWRMGFCSGIENSRHLTPQSTPTTLIDYFPDFMIVDESHVTIIPVQGMNGQARKOVVDGFLPSALDRPLRFEFEKHMNIYVSATGPEIETHDEMVEQIIRPTGLDPLDVPIEGQIDDLIGEIQARIERNERVLVTTLTKMSEDLTLYLKEIGIKVNLHSEIKTLERIEIIRDLRGKVDVGLNLREGLDIPEVSLVAILADKRGPLRSRSLIQTIGRAARNAEGVINYADTKTSMETAINETKRREQQRNEHEGTTPTKINKEIRDVIRATVYADKAEYITKAAPKLSKTKKERQKVEQMEHEKKEKALDFERAELDLLELKAG"

/complement(5166. .5396)

/gene="csbA"

/locus_tag="BSU35180"

/complement(5166. .5396)

/gene="csbA"

/locus_tag="BSU35180"

/function="unknown"

/codon_start=1

/transl_table=11

/product="putative membrane protein"

/protein_id="CAB15535.1"

/db_xref="GI:2636044"

/db_xref="GOA:P37953"

/db_xref="Subtilist:BG10501"

/db_xref="UniProt/Swiss-Prot:P37953"

/translation="MITKAVPALFFPPMLVLLFRVYFNNHYVALTRALLFASYLKGYTEYFVGLDVSLVAGLVMKKAKEKEE"

/5843. .8338

/gene="yvkC"

/locus_tag="BSU35190"

/5843. .8338

/gene="yvkC"

/locus_tag="BSU35190"

/function="unknown"

/note="similar to pyruvate, water dikinase"

/codon_start=1

/transl_table=11

/protein_id="CAB15536.1"

/db_xref="GI:2636045"

/db_xref="GOA:O34796"

/db_xref="InterPro:IPR002192"

/db_xref="InterPro:IPR008279"

/db_xref="Subtilist:BG14114"

/db_xref="UniProt/TrEMBL:O34796"

/translation="WISVLVRQAEESQLAGAKAMNLIKTKGLPYPDGFIOTNALARFEMDQLOETSNVEGGIISGFSELDELDTSSYYKLRBESTRSVAVSSASEDL EGAFPAQCYEYVNLKITEEFLAKVKECMSFSGRSYKKKANNQIABPLMGI VVOGLISESGYVFSNPPTHDRRELLISASGLGAVSVGVTPTPTVNSGSEIOKEIGAKIYMESAABEGIAEKTESEDMRSPCLTDEQVILAEITKTEKTLGYGVPDIENG IADQIYULQKRPITTIIDQKKAEEKRSMTITDNDNDWMKNESNIEGVPSPFSFIVALEYGLKRSQKRPFGVVDVEVLVYGHYISKQGGQPPSEDCGKELFPITLSEHMVDIINHYYLPFYRTLDOLAQTEHTESLSLEAQKLKATVLTAYEEHFIVFPQILLTNKLQAMVODIQGSENAHFYEMLTGRNKSLETRDCLMTFSVEVQENPNLLAFENN

KPEOLEKLEOTDEGRHFLKNVHEFLQEVGRSVKSHDLIIOIWEVNPYALANTONY
VRNGHFNPEROKTEKEXEKLYNELESIEBPGALREEDRYOQMTNANSNKDHHFY
IDAMLDARAFIPLKIGELLENVGIQUREDIWFLYDEVEDALHPVLSQKAEKR
QIFHEVELAOAPAYIGTPTKEQKAAEIVGAVIDEKENTENHIFGLAASSGIATGPV
KIRIDANESQAPGDLVCKMTPTPLMTSLFQDAKAILITDQGLISHAAILAREVGI
AVLGRTRATERLRDGDITVDSSESKITVSRSS"
8414..8983

gene
/gene="yvkB"
/locus_tag="BSU5200"
8414..8983
/gene="yvkB"
/locus_tag="BSU5200"
/function="unknown"
/note="similar to transcriptional regulator (TetR/AcrR family)"
/codon_start=1
/protein_id="CAB15537.1"
/db_xref="GI:2636046"
/db_xref="GOA:O34643"
/db_xref="InterPro:IPR001647"
/db_xref="Subtilis:BG1413"
/db_xref="UniProt/TREMBL:O34643"

Query Match 98.8%; Score 442.2; DB 1; Length 200690;
Best Local Similarity 99.3%; Pred. No. 3.5e-104;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATGTCGATCGATTTATATACATCGCATATTTAGAGTACTACTAGTTAATTTT 60
89478 ATGTCGATCGATTTATATACATCGCATATTTAGAGTACTACTAGTTAATTTT 89419
61 GCGGAAAAACAGGAGATGTCGCCGAGAGACTTGTGTACCGGATATTAGACTGTG 120
89418 GCGGAAAAACAGGAGATGTCGCCGAGAGACTTGTGTACCGGATATTAGACTGTG 89359
121 TTTAATCAGCCGCTTTATTTTACTGTTTCTAGTGAAGCTTCTCACTTATGTTATC 180
89358 TTTAATCAGCCGCTTTATTTTACTGTTTCTAGTGAAGCTTCTCACTTATGTTATC 89299
181 GTGAATACGGTATTACCAATTTATGATTTGTACGAGCAGGAAAAATTCGTCGCATG 240
89298 GTGAATACGGTATTACCAATTTATGATTTGTACGAGCAGGAAAAATTCGTCGCATG 89239
241 CTGATACAGGAGATGTCCTTAAATCGCTTGAATTTTCTATACCGGATTTACCATTT 300
89238 CTGATACAGGAGATGTCCTTAAATCGCTTGAATTTTCTATACCGGATTTACCATTT 89179
301 GAAATCGCAATTTTCGAGAAATCGGCATCATGTCGACGTTTAATGCCAATACCATTT 360
89178 GAAATCGCAATTTTCGAGAAATCGGCATCATGTCGACGTTTAATGCCAATACCATTT 89119
361 CAGAAACAGGTTTAAACATTAAGTTCGGAAGACGCTGTATTTGAGCGGAGGACCTTT 420
89118 CAGAAACAGGTTTAAACATTAAGTTCGGAAGACGCTGTATTTGAGCGGAGGACCTTT 89059
421 GCTATCATGTTTGTACTACTTAAT 447
89058 GCTATCATGTTTGTACTACTTAAT 89032

RESULT 6
E50426 450 bp DNA linear PAT 31-JAN-2002
LOCUS
DEFINITION Process for producing poly-gamma-glutamic acid.
ACCESSION E50426
VERSION E50426.1 GI:18629414
KEYWORDS JP 2001017182-A/3.
SOURCE
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 450)
AUTHORS Ashiuchi, M., Misono, H. and Soda, K.

TITLE
JOURNAL
Process for producing poly-gamma-glutamic acid
Patent: JP 2001017182-A 3 23-JAN-2001;
NAGASE ECO LTD

OS
Bacillus subtilis (hay bacillus) IFO 3336

PN
JP 2001017182-A/3

PD 23-JAN-2001

PF 09-JUL-1999 JP 1999196335

PR
MAKOTO ASHUCHI, HARUO MISONO, KENJI SODA

PI
C12N15/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC

PC
C12N9/00, C12N9/90, C12P13/14, C12N15/00, C12N5/00

CC
Key Location/Qualifiers
FT source 1.450
FT IFO 3336
FT /organism="Bacillus subtilis (hay bacillus)"

FEATURES
source
1.450
Location/Qualifiers
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/db_xref="taxon:1423"

ORIGIN

Query Match 98.6%; Score 440.6; DB 6; Length 450;
Best Local Similarity 99.1%; Pred. No. 2e-103;
Matches 443; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGTCGATCGATTTATATACATCGCATATTTAGAGTACTACTAGTTAATTTT 60
1 ATGTCGATCGATTTATATACATCGCATATTTAGAGTACTACTAGTTAATTTT 60
61 GCGGAAAAACAGGAGATGTCGCCGAGAGACTTGTGTACCGGATATTAGACTGTG 120
61 GCGGAAAAACAGGAGATGTCGCCGAGAGACTTGTGTACCGGATATTAGACTGTG 120
121 TTTAATCAGCCGCTTTATTTTACTGTTTCTAGTGAAGCTTCTCACTTATGTTATC 180
121 TTTAATCAGCCGCTTTATTTTACTGTTTCTAGTGAAGCTTCTCACTTATGTTATC 180
181 GTGAATACGGTATTACCAATTTATGATTTGTACGAGCAGGAAAAATTCGTCGCATG 240
181 GTGAATACGGTATTACCAATTTATGATTTGTACGAGCAGGAAAAATTCGTCGCATG 240
241 CTGATACAGGAGATGTCCTTAAATCGCTTGAATTTTCTATACCGGATTTACCATTT 300
241 CTGATACAGGAGATGTCCTTAAATCGCTTGAATTTTCTATACCGGATTTACCATTT 300
301 GAAATCGCAATTTTCGAGAAATCGGCATCATGTCGACGTTTAATGCCAATACCATTT 360
301 GAAATCGCAATTTTCGAGAAATCGGCATCATGTCGACGTTTAATGCCAATACCATTT 360
361 CAGAAACAGGTTTAAACATTAAGTTCGGAAGACGCTGTATTTGAGCGGAGGACCTTT 420
361 CAGAAACAGGTTTAAACATTAAGTTCGGAAGACGCTGTATTTGAGCGGAGGACCTTT 420
421 GCTATCATGTTTGTACTACTTAAT 447
421 GCTATCATGTTTGTACTACTTAAT 447

RESULT 7
E50424 3045 bp DNA linear PAT 31-JAN-2002
LOCUS
DEFINITION Process for producing poly-gamma-glutamic acid.
ACCESSION E50424
VERSION E50424.1 GI:18629412
KEYWORDS JP 2001017182-A/1.
SOURCE
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 3045)
AUTHORS Ashiuchi, M., Misono, H. and Soda, K.

TITLE Process for producing poly-gamma-glutamic acid
JOURNAL Patent: JP 2001017182-A 1 23-JAN-2001;
NAGASE ACO LTD
COMMENT OS Bacillus subtilis (hay bacillus) IPO 3336
PN JP 2001017182-A/1
PD 23-JAN-2001
PR 09-JUL-1999 JP 1999196335
PI MAKOTO ASHICHI, HARUO MISONO, KENJI SODA
PC C12N5/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
C12N9/00, C12N9/90,
PC C12P13/14, C12N5/00, C12N5/00
CC
FH Key Location/Qualifiers
FT source 1..3047
FT /organism='Bacillus subtilis (hay bacillus)
FT IPO 3336',
FT Location/Qualifiers
1..3045
/organism='Bacillus subtilis'
/mol_type='genomic DNA'
/db_xref='taxon:1423'
ORIGIN
Query Match 98.6%; Score 440.6; DB 6; Length 3045;
Best Local Similarity 99.1%; Pred. No. 1.6e-103;
Matches 443; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 ATGTTCCGATCAGATTATATACATCCACTAATTTAGGTGATCTCAGTTAATTTT 60
DB 1215 ATGTTCCGATCAGATTATATACATCCACTAATTTAGGTGATCTCAGTTAATTTT 1274
OY 61 GCGGAAAAACAGGGATGTCGCGGAGGAGACTGTGTGACCGGGATTTTATAGACTTGG 120
DB 1275 GCGGAAAAACAGGGATGTCGCGGAGGAGACTGTGTGACCGGGATTTTATAGACTTGG 1334
OY 121 TTTATACAGCCGCTTTATTTTACTTGTGTTGCTAGTAGCTTCTCATTATGTTATC 180
DB 1335 TTTATACAGCCGCTTTATTTTACTTGTGTTGCTAGTAGCTTCTCATTATGTTATC 1394
OY 181 GTGAATACGGTTTATCCAAATTTATGATTTTGTACGACGCAAAAATTCGTCGCCATG 240
DB 1395 GTGAATACGGTTTATCCAAATTTATGATTTTGTACGACGCAAAAATTCGTCGCCATG 1454
OY 241 CTGATACAGGGATGTCCTTAAATTCGGCTTGAATTTCTATACCCGATTGTACATTT 300
DB 1455 CTGATACAGGGATGTCCTTAAATTCGGCTTGAATTTCTATACCCGATTGTACATTT 1514
OY 301 GAAATCGAGAAATTTGAGGAATCGGATCATCTGCCAGTTTATTCGCAATACCAAT 360
DB 1515 GAAATCGAGAAATTTGAGGAATTCGGATCATCTGCCAGTTTATTCGCAATACCAAT 1574
OY 361 CAGAAACAGGTTTAAACCTTACGTTCCGAGACACGCTGCTATTTAGCGGAGCGACTTT 420
DB 1575 CAGAAACAGGTTTAAACCTTACGTTCCGAGACACGCTGCTATTTAGCGGAGCGACTTT 1634
OY 421 GCTATCATGTTTGTACTACTTAAT 447
DB 1635 GCTATCATGTTTGTACTACTTAAT 1661
RESULT 8
AB046355 3738 bp DNA linear BCT 11-JAN-2002
LOCUS AB046355
DEFINITION Bacillus subtilis ywsc, ywca, ywcb, ywcc genes, complete cds.
ACCESSION AB046355
VERSION AB046355.1 GI:13591556
KEYWORDS
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
REFERENCE 1
AUTHORS Urushidate, Y., Tokuyama, S. and Tahara, Y.

TITLE Characterization of the Bacillus subtilis ywsc gene, involved in
JOURNAL gamma-polyglutamic acid production
J. Bacteriol. 184 (2), 337-343 (2002)
PUBMED 11751809
REFERENCE 2 (bases 1 to 3738)
AUTHORS Tahara, Y. and Urushidate, Y.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2000) Yasutaka Tahara, Shizuoka University,
Faculty of Agriculture, 836 Ohya, Shizuoka city, Shizuoka 422-8529,
Japan (E-mail: acyrtah@agr.shizuoka.ac.jp,
Tel: 81-54-238-4878 (ex. 7808), Fax: 81-54-237-3028)
FEATURES
source Location/Qualifiers
1..3738
/organism='Bacillus subtilis'
/mol_type='genomic DNA'
/db_xref='taxon:1423'
gene 497..1678
/gene='ywsc'
CDS 497..1678
/gene='ywsc'
/codon_start=1
/transl_table=1
/protein_id='BAB40948.1'
/db_xref='GI:13591557'
/translation='MMLLIACAVILVIGLEKRRHQKIDALPVRVNINGIRKSTV
TRLTGLIAGYKVTGTTGDAPMIMPTPEKRIKRPQGNIGQKQVARETVE
RGAAIVSEGAIVPDYQIIFQELQANIVYVNEIDMDVNGPLDIAEFTAT
IPVNGHVIDSEYTERFKOKAKERNKVI IANRSKITDYLKFEFVPPDNASLIL
GVQALQIDETAFKMLNPPDGMARILPLSPSPRGFVNGFANASSTLNTWK
RVEIGYPTDDPIITNRCRADRVDTQFANDVPIEASELLIGQSEPIYAYAE
SKPADKLHOLEYKSTDEIMELKKRMHNVIVGVNIGHAEPLEKIHIEYKQOLV
S'
1693..2142
/gene='ywca'
CDS 1693..2142
/gene='ywca'
/codon_start=1
/transl_table=1
/protein_id='BAB40949.1'
/db_xref='GI:13591558'
/translation='MFGSDIYALILGLVLSLFAEKTGIVPAGLVYVGYLGVNPP
VFILVLVSLVILIVKYGSKFMILYGRKTPAMLTIGLIALDFLPIYFPI
AEFGIIGIVGGLANTVIOKGLITFGSTLLSGRAFAIMFYVYL'
2161..3303
/gene='ywcb'
CDS 2161..3303
/gene='ywcb'
/codon_start=1
/transl_table=1
/protein_id='BAB40950.1'
/db_xref='GI:13591559'
/translation='MKKELSFHEKLTQKQKKTNKHVFIPIVFLVFAFMAG
KAETPKYKTSVDVLSASFVGDIMMGVYKVTBOKADSIPOVEIIFPASDVAN
FENVITYOKNKKQADKEIHLQTNKESKVKLDNMFTVLSANNHAMDVIGQKMDTIG
EPATKNDIIVAGAGSLDAKKISYQVNGVLTATLGTFSVSGFPAKNTPGVLA
DPBIFPMISEAKGADIIVVQSHMGQSEYNDPDROROLRAMSDGADIIVGHNH
VLPEIYVNGTVIFYSLGNFVFDGQWRTSDSLVOYHLKNGGREGVTPIDHEAT
PAPYKDSLKQKTIIRLTQDSNPKMVBQKLTFLDIDHDKLSK'
3318..3485
/gene='ywcc'
CDS 3318..3485
/gene='ywcc'
/codon_start=1
/transl_table=1
/protein_id='BAB40951.1'
/db_xref='GI:13591560'
/translation='MKFVKAIWPFVAIVFMFMSAFKFNDDLTDEKOKIDWEMNKI
QQQSEPVNANK'
ORIGIN
Query Match 97.5%; Score 435.8; DB 1; Length 3738;
Best Local Similarity 98.4%; Pred. No. 2.7e-102;
Matches 440; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGTTGGATCAGATTATACATCGCACTAATTTAGTGTACTACTGATTATTTT 60
DB 1693 ATGTTGGATCAGATTATACATCGCACTAATTTAGTGTACTACTGATTATTTT 1752
QY 61 GCGGAAAAACAGGGATGTCGCGGACGAGCTGTGTACCGGGATTTAGACTGTG 120
DB 1753 GCGGAAAAACAGGGATGTCGCGGACGAGCTGTGTACCGGGATTTAGACTGTG 1812
QY 121 TTTAATCAGCCGCTTTATTTTACTGTTTGTCTAGTAGCTTCACTTATGTATC 180
DB 1813 TTTAATCAGCCGCTTTATTTTACTGTTTGTCTAGTAGCTTCACTTATGTATC 1872
QY 181 GTGAATAACGGTTTATCCAAATTTATGATTTGTACGAGCAAGAAATTCCTGCCATG 240
DB 1873 GTGAATAACGGTTTATCCAAATTTATGATTTGTACGAGCAAGAAATTCCTGCCATG 1932
QY 241 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGTATCTATACCCGATGTACCATTT 300
DB 1933 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGTATCTATACCCGATGTACCATTT 1992
QY 301 GAAATCGCAGATTTTCAGAGAAATCGGCATCATCTGTCAGGTTTATTCGCAATACCAT 360
DB 1993 GAAATCGCAGATTTTCAGAGAAATCGGCATCATCTGTCAGGTTTATTCGCAATACCAT 2052
QY 361 CAGAAACAAGGTTTACCAATTAAGTTTGTACGAGCAAGCTGCTATTTAGCGGAGCATTT 420
DB 2053 CAGAAACAAGGTTTACCAATTAAGTTTGTACGAGCAAGCTGCTATTTAGCGGAGCATTT 2112
QY 421 GCTATCATGTTTGTACTACTTAAT 447
DB 2113 GCTATCATGTTTGTACTACTTAAT 2139

RESULT 9
AB017333_36/c
WPCOMMENT

Sequence split into 43 fragments LOCUS AB017333 Accession AB017333

Fragment Name	Begin	End
AB017333_00	1	110000
AB017333_01	100001	210000
AB017333_02	200001	310000
AB017333_03	300001	410000
AB017333_04	400001	510000
AB017333_05	500001	610000
AB017333_06	600001	710000
AB017333_07	700001	810000
AB017333_08	800001	910000
AB017333_09	900001	1010000
AB017333_10	1000001	1110000
AB017333_11	1100001	1210000
AB017333_12	1200001	1310000
AB017333_13	1300001	1410000
AB017333_14	1400001	1510000
AB017333_15	1500001	1610000
AB017333_16	1600001	1710000
AB017333_17	1700001	1810000
AB017333_18	1800001	1910000
AB017333_19	1900001	2010000
AB017333_20	2000001	2110000
AB017333_21	2100001	2210000
AB017333_22	2200001	2310000
AB017333_23	2300001	2410000
AB017333_24	2400001	2510000
AB017333_25	2500001	2610000
AB017333_26	2600001	2710000
AB017333_27	2700001	2810000
AB017333_28	2800001	2910000
AB017333_29	2900001	3010000
AB017333_30	3000001	3110000
AB017333_31	3100001	3210000
AB017333_32	3200001	3310000
AB017333_33	3300001	3410000

AB017333_34 3400001 3510000
AB017333_35 3500001 3610000
AB017333_36 3600001 3710000
AB017333_37 3700001 3810000
AB017333_38 3800001 3910000
AB017333_39 3900001 4010000
AB017333_40 4000001 4110000
AB017333_41 4100001 4210000
AB017333_42 4200001 4225645
Continuation (37 of 43) of AB017333 from base 3600001 (AB017333 Bacillus licheniformis DC

Query Match 65.1%; Score 290.8; DB 1; Length 110000;
Best Local Similarity 78.3%; Pred. No. 6.8e-65;
Matches 349; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1 ATGTTGGATCAGATTATACATCGCACTAATTTAGTGTACTACTGATTATTTT 60
DB 6638 ATGTTGGATCAGATTATACATCGCACTAATTTAGTGTACTACTGATTATTTT 66329
QY 61 GCGGAAAAACAGGGATGTCGCGGACGAGCTGTGTACCGGAGTATTTAGACTGTG 120
DB 66328 GCGGAAAAACAGGGATGTCGCGGACGAGCTGTGTACCGGAGTATTTAGACTGTG 66269
QY 121 TTTAATCAGCCGCTTTATTTTACTGTTTGTCTAGTAGCTTCACTTATGTATC 180
DB 66268 TTTAATCAGCCGCTTTATTTTACTGTTTGTCTAGTAGCTTCACTTATGTATC 66209
QY 181 GTGAATAACGGTTTATCCAAATTTATGATTTGTACGAGCAAGAAATTCCTGCCATG 240
DB 66208 GTGAATAACGGTTTATCCAAATTTATGATTTGTACGAGCAAGAAATTCCTGCCATG 66149
QY 241 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGTATCTATACCCGATGTACCATTT 300
DB 66148 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGTATCTATACCCGATGTACCATTT 66089
QY 301 GAAATCGCAGATTTTCAGAGAAATCGGCATCATCTGTCAGGTTTATTCGCAATACCAT 360
DB 66088 GAAATCGCAGATTTTCAGAGAAATCGGCATCATCTGTCAGGTTTATTCGCAATACCAT 66029
QY 361 CAGAAACAAGGTTTACCAATTAAGTTTGTACGAGCAAGCTGCTATTTAGCGGAGCATTT 420
DB 66028 CAGAAACAAGGTTTACCAATTAAGTTTGTACGAGCAAGCTGCTATTTAGCGGAGCATTT 65969
QY 421 GCTATCATGTTTGTACTACTTAAT 446
DB 65968 GCTATCATGTTTGTACTACTTAAT 65943

RESULT 10
CP000002_36/c
WPCOMMENT

Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000
CP000002_11	1100001	1210000
CP000002_12	1200001	1310000
CP000002_13	1300001	1410000
CP000002_14	1400001	1510000
CP000002_15	1500001	1610000
CP000002_16	1600001	1710000
CP000002_17	1700001	1810000
CP000002_18	1800001	1910000
CP000002_19	1900001	2010000

CP000002_20 2000001 2110000
CP000002_21 2100001 2210000
CP000002_22 2200001 2310000
CP000002_23 2300001 2410000
CP000002_24 2400001 2510000
CP000002_25 2500001 2610000
CP000002_26 2600001 2710000
CP000002_27 2700001 2810000
CP000002_28 2800001 2910000
CP000002_29 2900001 3010000
CP000002_30 3000001 3110000
CP000002_31 3100001 3210000
CP000002_32 3200001 3310000
CP000002_33 3300001 3410000
CP000002_34 3400001 3510000
CP000002_35 3500001 3610000
CP000002_36 3600001 3710000
CP000002_37 3700001 3810000
CP000002_38 3800001 3910000
CP000002_39 3900001 4010000
CP000002_40 4000001 4110000
CP000002_41 4100001 4210000
CP000002_42 4200001 4222334
Continuation (37 of 43) of CP000002 From base 3600001 (CP000002 Bacillus licheniformis A

Query Match 65.1%; Score 290.8; DB 1; Length 110000;
Best Local Similarity 78.3%; Pred. No. 6.8e-65;
Matches 349; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 1 ATGTTGGATCAGATTATATACATGCACTAATTTAGGTACTACTAGATTATTTT 60
Db 66228 ATGTTGGATCAGATTATATATGCTCCCTCATTTTAGAGTCTTACAGTTGATTTT 66169
Qy 61 GCGGAAAAACAGGATCGTGCAGGACGCTGTTGACCGGATTTTAGAATTGTG 120
Db 66168 GCAGGAAAAACGGAAATGTACACAGCCGCTCGTACCGGATTATTTGGAGCTTGTC 66109
Qy 121 TTATATCAGCCGCTTATTTTCTTTGCTAGTAGCTTCTCACTATGTTATC 180
Db 66108 TTCAATCAGCCGATTTTCAATGCTGCTCTTTTGTCACTTCTCACTATGTTATC 66049
Qy 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGAACGACGAGAAAAATTCGCTGCCATG 240
Db 66048 GTGAAATTCGACCTTCCAAATTTATGATTTTGAACGACGAGAAAAATTCGACGAATG 65989
Qy 241 CTGATTAACAGGATCGTCTTAAAAATCGCGTTGATTTTCTATACCGAATGTACATT 300
Db 65988 CTGATTAACGGGAATCTTTGAAAAATCGGTTTGAATTTATATATCCGATGCGGTTT 65929
Qy 301 GAAATCGCAGATTTCAGAAATCGGCATCATGCGCAGGTTTAAATGCAATACCAT 360
Db 65928 GAGATTCGCGAATTCAGGGAATCGGAATCATGCGCGGCTATGCGCAATACCAT 65869
Qy 361 CAGAAACAGGTTTAAACATTACGTTCCGAGACAGCTGCTATTAGCGGAGCACTTT 420
Db 65868 CAAAGACAGGATTAACATTACGTTCCGAGAGTACGCTTTATTAGCGGAGCAACATTC 65809
Qy 421 GCTATCATGTTGTTTACTACTTAAT 446
Db 65808 GTCAATTATGATGCTTACTACTTAAT 65783

RESULT 11
AX434510 450 bp DNA linear PAT 28-JUN-2002
LOCUS Sequence 2925 From Patent WO0229113.
ACCESSION AX434510
VERSION AX434510.1 GI:21659318
KEYWORDS
SOURCE Bacillus licheniformis
ORGANISM Bacillus licheniformis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1

AUTHORS Berka, R. and Clausen, I. G.
TITLE Methods for monitoring multiple gene expression
JOURNAL Patent: WO 0229113-A 2925 11-Apr-2002;
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
source Location/Qualifiers
1..450
/organism="Bacillus licheniformis"
/mol_type="unassigned DNA"
/db_xref="taxon:1402"
ORIGIN
Query Match 64.7%; Score 289.2; DB 6; Length 450;
Best Local Similarity 78.0%; Pred. No. 3.6e-64;
Matches 348; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 1 ATGTTGGATCAGATTATATACATGCACTAATTTAGGTACTACTAGATTATTTT 60
Db 1 ATGTTGGATCAGATTATATATGCGCTCATTTTAGAGTCTTACTCAGTTGATTTT 60
Qy 61 GCGGAAAAACAGGATCGTGCAGGACGACTTGTGACCGGATATTAGACTTGTG 120
Db 61 GCAGGAAAAACGGAAATTTGACACCGGCTGCTGACCGGTTATTGGACTTGTG 120
Qy 121 TTATATCAGCCGCTTATTTTACTTGTGTTGCTAGAGCTTGTCACTTATGTTATC 180
Db 121 TTCAATCAGCCGATTTTCAATGCTGCTGCTTTTGTGCTAGTTGCTGAGTATGATC 180
Qy 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGAACGACGAGAAAAATTCGCTGCCATG 240
Db 181 GTGAAATTCGACTTCCAAATTTATGATTTTGAACGACGAGAAAAATTCGACGAATG 240
Qy 241 CTGATTAACAGGATCGTCTTAAAAATCGGTTGATTTTCTATACCGAATTTGACATT 300
Db 241 CTGATTAACGGAATCTTTTGAATATGCTTTGATTTTATATATCCGAGATCCGTT 300
Qy 301 GAAATCGCAGATTTCAGAAATGCGATCATGCTGCAGGTTTAAATGCAATACCAT 360
Db 301 GAGATTCGCGAATTCAGGGAATCGAATCATGCTGCGGCGTATGCGCAATACCAT 360
Qy 361 CAGAAACAGGTTTAAACATTACGTTGGAAGACGCTGCTATTTAGCGGAGCACTTT 420
Db 361 CAAAGACAGGATTAACATTACGTTGGAAGTACGCTTTTATTTAGCGGAGCAACATTC 420
Qy 421 GCTATCATGTTGTTTACTACTTAAT 446
Db 421 GTCAATTATGATGCTTACTACTTAAT 446

RESULT 12
BACCAPABC 3244 bp DNA linear BCT 26-APR-1993
LOCUS BACCAPABC
DEFINITION B.anthraxis encapsulation protein genes (capA, capB, and capC),
complete cds.
ACCESSION M24150
VERSION M24150.1 GI:142630
KEYWORDS encapsulation protein; membrane-associated protein.
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.
REFERENCE 1 (bases 1 to 3244)
Makino, S., Uchida, I., Terakado, N., Sasakawa, C. and Yoshikawa, M.
Molecular characterization and protein analysis of the cap region,
which is essential for encapsulation in Bacillus anthracis
J. Bacteriol. 171 (2), 722-730 (1989)
JOURNAL PDBMED
COMMENT Original source text: B.anthraxis (strain TE702; isolate PCAP1)
DNA.
Draft entry and computer-readable sequence for [1] kindly provided
by I.Uchida, 01-MAY-1999.
FEATURES
source Location/Qualifiers
1..3244
/organism="Bacillus anthracis"

CDS
/mol_type="genomic DNA"
/db_xref="taxon:1392"
282..1475
/note="44 Kd encapsulation protein CapB"
/codon_start=1
/transl_table=11
/protein_id="AA02286.1"
/db_xref="GI:142631"
/translation="MFIIGCTVFLIYIGWEORCHOKRLNSIPRNINIGRKST
VTRLITGVVOEAKYTVGTGTSTGTSARMTYMFDEBQPIKRREKGNIGQRVVEAA
DLEALALICEGAQVDPQIIFQNKQIQAQYIVNVEDHMDVAGPLIDEVAFEA
TTPVNGHVTTESELDYKEVAERNTKIVADSRISSEELREFEDVFPDNLASLA
LAVAFALGIDETAFRGMNAPDGRNIRIFADQSPAPFANGPANDPESLRITW
EKVDGFSNLAIPVIMNCRPDRVDRTEQPADVPIYKAEIVIAIGETAPITSAE
KGDIPTOEYWNLEGSTSEINSMRMPYLKRLIVYGVNTHGAEPILDMIEQIGKK
QAKVI"

CDS
1490..1939
/note="16 Kd encapsulation protein CapA"
/codon_start=1
/transl_table=11
/protein_id="AA02287.1"
/db_xref="GI:142632"
/translation="MGSGLYIALVIGVTLIFTERTGILPAGLVPEGLAVENOP
VPMVAVRSLITLVIVYGVSRPMLVGRREKPAATLIGICLKLFDVCPVMPFEI
FEPRLGIVPEGLIANTIORGLPLITGTTILSGATPAINMIVYLF"
1951..3186
/note="46 Kd encapsulation protein CapA"
/codon_start=1
/transl_table=11
/protein_id="AA02288.1"
/db_xref="GI:142633"
/translation="MKRKLTFQEKLLIFIKTKKNPRVAVLPLIALVILIAATWQ
RLEAVAPVGRHRENEKLTMTVDIMWGRHKEIVNRVGYDVFRRHSPYLKSDVYSG
NEHEPVLLEDKNYQKADKNIHLSAKEEYKAVKEGFTVLMANNHMDYAKGTGD
TKAFKEDLDVYAGENFKDYNIYQVNGVRAVTLGTDPAFVGAIALTEQPSLT
SNRPVULKOISKADPKRKNADIVVNNHMEYDNRSPROBALAKAMVADGDI
VGHHPVLOSFDYKOGIIFYSIGNVPVQGWRTQDSALVOYHARDNTALIDVYPL
NTOEGSPKPVASALDKNRVRYRQUTKTSKGLMSKDDKLEIKLNKHVIEKKREK
QEHODKQEKENOVSEVET"

CDS
/note="46 Kd encapsulation protein CapA"
/codon_start=1
/transl_table=11
/protein_id="AA02288.1"
/db_xref="GI:142633"
/translation="MKRKLTFQEKLLIFIKTKKNPRVAVLPLIALVILIAATWQ
RLEAVAPVGRHRENEKLTMTVDIMWGRHKEIVNRVGYDVFRRHSPYLKSDVYSG
NEHEPVLLEDKNYQKADKNIHLSAKEEYKAVKEGFTVLMANNHMDYAKGTGD
TKAFKEDLDVYAGENFKDYNIYQVNGVRAVTLGTDPAFVGAIALTEQPSLT
SNRPVULKOISKADPKRKNADIVVNNHMEYDNRSPROBALAKAMVADGDI
VGHHPVLOSFDYKOGIIFYSIGNVPVQGWRTQDSALVOYHARDNTALIDVYPL
NTOEGSPKPVASALDKNRVRYRQUTKTSKGLMSKDDKLEIKLNKHVIEKKREK
QEHODKQEKENOVSEVET"

ORIGIN 748 bp upstream of HindIII site.

Query Match 53.1%; Score 237.4; DB 1; Length 3244;
Best Local Similarity 70.7%; Pred. No. 7.5e-51;
Matches 316; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

1 AGTTGGATCGATTATATCATGCGACATATTTAGTGTACTCATGTTAATTTT 60
1490 ATGTTGGATCGATTATATCATGCGACATATTTAGTGTACTCATGAGCCTTATTT 1549
61 GCGAAGAAAAACAGGATCGTCCGCGAGCATTTGTTACCGGATATTTAGGACTGTG 120
1550 ACGAAGAAACAGGATTTTACTCGGAGGTTAGTTAGTTACTCGTTATTAGACACGTT 1609
121 TTTAATCAGCGGCTTTATTTACTGTTTTCGTAGTACGCTTGCCTATTATGTTATC 180
1610 TTTAATCAGCGGCTTTATTTACTGTTTTCGTAGTACGCTTGCCTATTATGTTATC 1669
181 GTGAATPAGCGTTATTCAGAAATTTATGATTTTACGACGACGAGAAATTCGCTGCATG 240
1670 GTTACGTATGCTTTTCAGAAATTCATGATTTTATATGCGCTGAGAAATTTGCGGCAAG 1729
241 CTGATTAACAGGATCGTCTTAAATCGCTTGATTTTCTATPACCGGATTTGACCAATTT 300
1730 CTAAATACAGGATTTGTTTAAACCTTTATTTGATATGTTATCTCTGTATGACCAATTT 1789
301 GAAATGACAGATTTGAGGAATCGCATCATGCTGCAGTTAAATTTGCCATATCATTT 360
1790 GAGATTTTAAATTCCTGTATTTGAGGATTAATTTGTTCCAGATTAATTTGCAAAATPACATTT 1849
361 CAGAAACAGGTTTAACCATTTAGCTTCGAGAACGCGTGCATTTAGCGAGGACGACCTTT 420
1850 CAAGACAGGTTTACCATTTAACCATTTGAACTACATTTTGTATGTTGCTGCAACATTT 1909

Qy 421 GCTATCATGTTGTTACTACTTAATT 447
DB 1910 GCAATCATGATATTTATTTACTTAATT 1936

RESULT 13
AE011191/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacillus anthracis str. A2012
Bacillus anthracis str. A2012
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.

REFERENCE
AUTHORS
1 (bases 1 to 94829)
Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umeyam, L., Jiang, L.,
Holtzapple, E., Busch, J.D., Smith, K.L., Schnupp, J.M., Solomon, D.,
Keim, P. and Fraser, C.M.
Comparative genome sequencing for discovery of novel polymorphisms
in Bacillus anthracis
Science 296 (5575), 2028-2033 (2002)

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
2 (bases 1 to 94829)
Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umeyam, L., Jiang, L.,
Holtzapple, E., Busch, J.D., Smith, K.L., Schnupp, J.M., Solomon, D.,
Keim, P. and Fraser, C.M.
Direct Submission
Submitted (09-MAY-2002) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA

FEATURES
source
1..94829
/organism="Bacillus anthracis str. A2012"
/mol_type="genomic DNA"
/strain="A2012"
/db_xref="taxon:191218"
/plasmid="pX02"
/note="Florida strain"
complement (543..1139)
/gene="BxB0001"
complement (543..1139)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-03)"
/protein_id="AA026162.1"
/db_xref="GI:20520281"
/translation="MTYKTLFQFRNGQIHVDIVDPOGQYPEYPRPADVLRAHATVLD
EYMRGKGQDDVSLKREYKDKOSVNTKQMDVLIKLMELIAQSGNSLSLEKLEQ
EYKSEIASVSTKSGKFPRTYKREIENYVGHRIQREVTTSFVKPLPMTIDV
QIVMKKGKHQVPTRLGTGYQEIEMEQIPEHRRKLE"
complement (1151..1438)
/gene="BxB0002"
complement (1151..1438)
/gene="BxB0002"
complement (1151..1438)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-04)"
/protein_id="AA026163.1"
/db_xref="GI:20520282"
/translation="MEIHVRNVPYHLKIDKCEIKGKIGRRYRYMEYINMPEQH
FNOEYRNKEDKEDFDEAVSNVSIITLDROSDKQOEYIDATHELVAAAIKLNKE"
complement (1652..1768)
/gene="BxB0003"
complement (1652..1768)
/gene="BxB0003"
complement (1652..1768)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

gene
CDS
KGEHRNNLKITDKKVVIVAPNTDDLGDPNFKSHVKAALYINSEGVSKCTALLD
IEYIEGLENKQTINSVVKITLEKQGEIKVYENPVPVKR"
/complement (3589..4731)
/locus_tag="GBAA_PX02_0007"
/locus_tag="GBAA_PX02_0007"
/complement (3589..4731)
/locus_tag="GBAA_PX02_0007"
/note="identified by match to PFAM protein family HM
PF00877"
/codon_start=1
/transl_table=1
/product="NLP/P60 family protein, (px02-08)"
/protein_id="AAT8937.2"
/db_xref="GI:47552349"
/translation="MAATQETALIDYKKVKRIRKIVRLGGSTGVVIAAATLLIYS
MAIFGOSSTGTPNGISGTATVKNLPPEVRNOMAEQCAAGPELVVVALIM
VESNGISBEKLPDINQSSSGMAMNTISNPDSIYGMHAKGAFDAKMGINDILA
IVQYINRGNVHMLAANKKTHSIOTADYSLIVAPGGRNGTIGISQVAVAN
GGYRINGNFVFAEMVKQYISFDGAGTSGQIPGSETFFVMDEVLKYNVWVG
GKSSQGFDCGLTYMAVKTAGITIPISAATQYDFVEDEKDAQPGDLVFRGTGG
PNHVSIGIYIDANTMYDNGSGVGWQHQTSSYQOHVAGIRVRP"
/complement (4733..6667)
/locus_tag="GBAA_PX02_0008"
/complement (4733..6667)
/locus_tag="GBAA_PX02_0008"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (px02-09)"
/protein_id="AAT8938.2"
/db_xref="GI:47552350"
/translation="WVRLGKKEQKQKRTGYDLDFISHPQGGIKENEKXIAOGDCY
VACHVYSIAEDIPMLTALNNKDTISSVDVATANKEEYVKDINRITLKDQMS
ERBSTPDADAEVQNLTPDPAISITQOGEIVKVKLRIVYDPILEQERIGDITKE
IAGQYKAOVYVFKQKEWQTLFASDYQIYLGKSYRPSKNIIGIPFPHQDCK
DPRIYLGQISTGAFILDPFSTGTRTSFGFIPGKAGKSTLQKLEGLAKODC
FIRGDKARDVYVVOQGGKIDLSGVQDEEVEGTINLELVKATKITVTGAVDER
GSFIOHISKVNMIRFLNTEFDDVYIQEFGHLYAFYEMGLVQKSGDRPYKTYGP
ENHYPILDEPFYKISNKLBAEAGAPQHRDLAIKQVEDMITYGDMFNHTLLNK
ENBQIVPEIDIGISKYDKSVNFCOLFTALTILMGAHNGRQMYLREKRLSIEDYK
YRWFPLDECHAVINSQNEFANRYVIEFRENRKFPAGIFPRTQSPQENLPPRGSTVDY
STMAIFELTOYKIFLNNDNSVLETLKVLDSLETSERILPELKREALVQVSSTE
TYNWFDPDEKQLEKFGQ"
/complement (6684..7358)
/locus_tag="GBAA_PX02_0009"
/complement (6684..7358)
/locus_tag="GBAA_PX02_0009"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (px02-10)"
/protein_id="AAT8939.2"
/db_xref="GI:47552351"
/translation="WKMVAATBAKKNRKAQISEKNGKQSFISTIFASFKNITDAS
IVDIAVYRLEQKTKGILVDHVRNQLQVYLKVTTDLISNODLKRFNQQLSLRVYH
EPLKILSTVSTETEEQVYKRMALRQGSQSEVSEKHEHLYQCSIAINLNR
VLWVEKNLKELAFLIVYGNKETELIKVXKMKRYGGRFNLQNNKAKEVERKLFXKQ
NMNSEM"
/complement (7416..7730)
/locus_tag="GBAA_PX02_0010"
/complement (7416..7730)
/locus_tag="GBAA_PX02_0010"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (px02-11)"
/protein_id="AAT8940.2"
/db_xref="GI:47552352"
/translation="MEMMRNPKNTKQETIKLAFYIIDAATIALMLVLSVMPKVPVGS
GGRMFYVLTGTRFLPLCIKPHNSPTNRNIFVILDMKMNKNHPIEAVTISSEYK
RK"
/complement (7733..7936)
/locus_tag="GBAA_PX02_0011"
/complement (7733..7936)
gene
CDS

/locus_tag="GBAA_PX02_0011"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (px02-12)"
/protein_id="AAT8941.2"
/db_xref="GI:47552353"
/translation="MNMWIFAILMLGVAIVLSIAITFPFKOIFTFKRRKKKEPVWTV
TNAEYLDKDRFMKKEQKQKKEHD"
Query Match 53.1%; Score 237.4; DB 1; Length 94830;
Best Local Similarity 70.7%; Pred. No. 4,8e-51;
Matches 316; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
QY 1 ATGTTGGATCAGATTATATCATGACGACTAATTTAGTGTGACTACGATTATTTT 60
|||||
DB 55584 ATGTTGGATCAGATTATATATGCTATTGATGATATTAGAGATTACACGAGCTTATTTT 55525
|||||
QY 61 GCGGAAAAAACAGGATCTGCGCGGACGACCTGTTACCGGATATTTAGACTTGTG 120
|||||
DB 55524 ACAGAAAGAACAGGATTTTAACTGACGAGTTAGTTAGTACCTGGTATTTAGACATCGTT 55465
|||||
QY 121 TTTTATCAGCCGCTTTATTTTACTGTTTGTAGTACGCTTGCATTATGTTATC 180
|||||
DB 55464 TTTTATCAGCCGCTTTATTTAGTTGTTGTTTATTTATCATGATTTTAACTATGTTATC 55405
|||||
QY 181 GTGAATACGGTTTATCCAAATTTATGATTTTGTAGCGAGCAGAAATTCGCTGCATG 240
|||||
DB 55404 GTTACGTATGTTGTTTCAAGTTATATATTTATATGACCGGTAAATAATTTGCGGCAACG 55345
|||||
QY 241 CTGATTAACAGGATCGTCTTAAATATCGCTTGAATTTTCTATACCGGATTTGACCATTT 300
|||||
DB 55344 CTATTTACAGGATTTGTTTAAACCTTTATTTGATTTATGTTATCGGTTATGCAATTT 55285
|||||
QY 301 GAAATCGCAGATTTGAGGATCGCATCATCTGCGAGCTTTAATTTGCCAATACCAT 360
|||||
DB 55284 GAGATTTTGTGATTTCCGTTGATTTGAGTTATGTTTCAGAGATTAATTTGCAAAATACAT 55225
|||||
QY 361 CAGAAACAGGTTTAAACATTTAGTTGGAAGCAGCGTGCATTTAGCGGAGGACCTTT 420
|||||
DB 55224 CAAGACACAGGTTTACCATTTTACCATTTGGAACATTAATTTGTTAAGTGTGCAACATTT 55165
|||||
QY 421 GCTATCATGTTTGTCTTACTACTTAAAT 447
|||||
DB 55164 GCATTCATGATATTTATTTACTTAAATTT 55138
|||||
RESULT 15
AF188935/C
LOCUS 96231 bp DNA circular BCT 01-OCT-2003
DEFINITION Bacillus anthracis plasmid px02, complete sequence.
ACCESSION AF188935
VERSION AF188935.1 GI:6470151
KEYWORDS
SOURCE
ORGANISM
Bacillus anthracis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.
1 (bases 1 to 96231)
REFERENCE
AUTHORS Okinaka,R.T., Cloud,K., Hampton,O., Hill,K.K., Keim,P., Lamke,G.,
Kunano,S., Mauter,D., Martinez,Y., Svensson,R., Tatam,L.R.,
Brown,A.B. and Jackson,P.J.
TITLE Submitted (01-NOV-1999) BioScience Division, Los Alamos National
Laboratory, M888, Los Alamos, NM 87545, USA
FEATURES
source
1..96231
/organism="Bacillus anthracis"
/mol_type="genomic DNA"
/db_xref="taxon:1392"
/plasmid="px02"
complement (19..477)
/note="hypothetical protein"

/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-01"
/protein_id="AAFI3606.1"
/db_xref="GI:6470152"
/translation="MRGKGDDVSLKKEVKDLKOSVNTMKQMDVLIKIMELIAO
SGSNLSLEKLOEVKSEIASSVTKSEKPRITVYKIEENVOHQIIOEVKTTFS
VKPLPMITDVQIVMKKGKQVPTRLGTGYOIEWEQIPEHRKELE"
complement(482..1018)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-02"
/protein_id="AAFI3607.1"
/db_xref="GI:6470153"
/translation="MGKSPSITLTSQYITPNKKSFTYVYCMROKALLKDGHTLNT
ELTELSRQNALDSPDMERGSPHSQKVSSEKAEKAVLLTSPNSNTTDDPKYV
SNWKKYALDESKKKTKTEEBEVSILNKHITDLSPEMKEGKKEGIKAEIISGVFSK
DKNVITSTDNLNKIDITR"
complement(1037..1633)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-03"
/protein_id="AAFI3608.1"
/db_xref="GI:6470154"
/translation="MTYVKLRFQFGNOIHVDIVDFORQYREYPREADVLRHAYVLD
EKRRBGQDDVSLKKEVKDLKOSVNTMKQMDVLIKIMELIAOSSGNLSLEKLO
EVKSEIASSVTKSEKPRITVYKIEENVOHQIIOEVKTTFSVKKPLPMITDVQ
IVMKKGKQVPTRLGTGYOIEWEQIPEHRKELE"
complement(1645..1932)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-04"
/protein_id="AAFI3609.1"
/db_xref="GI:6470155"
/translation="MEIHRYANVPYHLKIDRCKEIGKKIGRRYVRMEYIMMPEQH
FQOYSRNKEDFDEAVSNVSTILDRQSKLOEYIDATIELVAAMIKLNEE"
complement(2240..3088)
/note="similar to Dictyostelium discoideum
glutamine-asparagine rich protein encoded by Genbank
Accession Number U07817"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-05"
/protein_id="AAFI3610.1"
/db_xref="GI:6470156"
/translation="MSIKFRMNKGRKPRNTRKAYILFTLGSPTMFEVLCISPVESG
ATYKPEMKIGEQSLSTYKIAVAKEINPDQTLRIYELPADDSQILSMKRY
ENKYIKOXNNVTKYVRASDNYIVISEVPEEFVSSVAVPEYIHELQNDVDL
KERSMAYVLENEKLIINRELKKSDFEYEREYAFSQOARKEIEKMEKESAMQOL
KIKNEQLTKMEYQTEGEKVTKNITINSNESTINNHOKEIDVLKEDIKKMEKKIQLLD
EKKTI"
complement(3085..3360)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-06"
/protein_id="AAFI3611.1"
/db_xref="GI:6470157"
/translation="MLSELKPNHDYAKESGYLILSRKKKGVKDKFLIPIPTWFDY
NPEKYEMLIVREXOPVNGEKYTNCKLENINHAQVSVVAVKGERMK"
complement(3449..4057)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental

/transl_table=11
/product="PX02-07"
/protein_id="AAFI3612.1"
/db_xref="GI:6470158"
/translation="MNNLSKLSGALYLSFNGCIWAANTTKEPDKDKRYRVYSEK
EKOIGDLKVLKLEQDKKINEGAGENKEGNTLIDLNKREVAHQFHAFLDYSVN
KGERNNILKTTDKKVVIVAPNTDDGDLNPKFSHVKAALYINSEGVSKCTALLD
IEYITGLENKQTTINSVVKITLEDQSEELKVEVSNPYVR"
complement(4083..5225)
/note="similar to Enterococcus faecalis transposon TN916
ORF4 encoded by Genbank Accession Number AB001488"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-08"
/protein_id="AAFI3613.1"
/db_xref="GI:6470159"
/translation="MAATQETAIDKRYKKRIKMIVRLLGSGTGVVIAATLLIIVS
MAIRGGQSTGTPNGGISGATVKNLPEVWRQWAMEQECAGQVPELVYVLAIIIM
VESNGISEKLBDIQOSSSGQAMANTISNPQDSIYGVWHLKGAFDKMGINDLLA
IVQYVNRGRVWHLNANKTHSIOADYSLIVVAPAGRNNGTTIGSQPVAVAN
GGIRYINGANFYYAMVQYLSPDAGTSGOIPGGSSETFVMDDEVLYKXGNPYVWG
GKSSQGFDCSGLTYMAIKTIGITIPISAATQIPTVEVDPKDAQPGLVFPKGTGG
PNVHSVGIYIDANTMYDSNGSGVGHQFTSSVQOHHYAGIRRVPR"
complement(5227..7158)
/note="similar to Staphylococcus aureus transfer complex
protein trsf encoded by Genbank Accession Number L11988"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-09"
/protein_id="AAFI3614.1"
/db_xref="GI:6470160"
/translation="MKLSGKQOQKRTGCIYLDPISRVQOGIKENKRYIAQGCYV
ACIHVYSLAEDIPLMLTALMNNKDTISSVVAATNKEVVKDINRSITELKDRITSE
RSTGRDADHMLONLTDPARSITQOGIIVKLYKRIYIYDVLVQIEKRGIDIKKEI
AGONKAQVVEKQKEWOTLPASDQIEYVAGSYGLPSKNIQELGFIPIHHQDLKD
PRGIYIGOTSGCAFIIDPFGSTGRSIFGKMGAGSTLLKQJEGVADGCF
IRGPKARDYTTVVOQSGKITIDLSGVQDSEVEVMGNINPEVATKLTNGGAVDEG
SFIQHSVYTNMIRPLNTEPDVTIVIQERKHLIYATITKGLVPQKGSRPYVYVCP
NHYPILDEPFYKYSMLKLEAGATPQRHDLKIQVMDMITVYDMEFNGITLXFE
NEQIVPEIDIGISKYDKSVNQLTALTALIMSHALKNGRQMKYLRBENKULSIEVYK
FMVPLDECHVYINSQNEFANRYLFESEMERKFPAGITQSPQEMLPPEPSITDVS
TMKAFELTQYKIFLPMNDSVLETIKSVLGSFLTSESRILPELRGEALVQVSTET
YXVMPDPBQKQERKQGO"
complement(7178..7846)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-10"
/protein_id="AAFI3615.1"
/db_xref="GI:6470161"
/translation="MKVATAEAKVNRKQKQISEKSNQSGFSTIFPASFNTDASIV
DIAPYKRLQOKTGLIVDHRNNLOVYLKVTYDILLNNODDLKRFNNQULSLGRVYHEP
PKILSTYSTETEOOVVYKRWALRFQSGQSEVSEKKEHLMYORYSLAINLRVL
WVKULKEIAFTVYVQKNETLILNKVDMKRYGGRQPLQMKAKVEYKLIPLXQNM
NSEM"
complement(7910..8224)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-11"
/protein_id="AAFI3616.1"
/db_xref="GI:6470162"
/translation="MEMMNPKNTOEIRGLAFYIIDGAILALMEVLASYMPKVPVG
GGRIMFYLFGTFLGFLICIKRNSPTNRNITVIIIDMLKMDKKNYHPIEIVNTISSTK
RK"
complement(8227..8430)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental

CDS

```
/transl_table=1
/product="PX02-12"
/protein_id="AAFI3617.1"
/db_xref="GI:6470163"
/translation="MNNMIFAILMLGVAIVLSIATFEKQIFTIKRRKKKEPWWTV
TNAEYLDIDFWKKEQDKKEHD"
complement(8432..8686)
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-13"
/protein_id="AAFI3618.1"
/db_xref="GI:6470164"
/translation="MKVIDANRRRIYFEMKQELRASILMTVAGFIIFAILVFOIS
FELGHLVHYIVFAFLYLSLHLYSNNKLARKIEKQGY"
complement(8704..11562)
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-14"
/protein_id="AAFI3619.1"
/db_xref="GI:6470165"
/translation="MATMKIKEMFPOSKKALVIGLAVITLFSFTFPIAVAFDWE
FLGKDVTEKVAFLONEMIYADFLGISWISOGGIIKTLIVLSALEGLIDPTFS
FPDIKDGNDPASIIRKGLFYALFVVIWIGRTIIOHKPRPKSVGNILMITG
LLGGLNELADMKRSTDPFSETTYSKAKDGLADLVKONTADILYISKTFEPIOS
KKENEPALSTDSKPKNGMSKDIPLKQGLGVVTPKVIEMLSQDIPKETEYLVYKI
TNDDEKTEVKINDSVFNPFPKDPFGYVRYPMNGTIFMGLALGVAVLFVVFVFM
```

CDS

```
/transl_table=1
/product="PX02-14"
/protein_id="AAFI3619.1"
/db_xref="GI:6470165"
/translation="MATMKIKEMFPOSKKALVIGLAVITLFSFTFPIAVAFDWE
FLGKDVTEKVAFLONEMIYADFLGISWISOGGIIKTLIVLSALEGLIDPTFS
FPDIKDGNDPASIIRKGLFYALFVVIWIGRTIIOHKPRPKSVGNILMITG
LLGGLNELADMKRSTDPFSETTYSKAKDGLADLVKONTADILYISKTFEPIOS
KKENEPALSTDSKPKNGMSKDIPLKQGLGVVTPKVIEMLSQDIPKETEYLVYKI
TNDDEKTEVKINDSVFNPFPKDPFGYVRYPMNGTIFMGLALGVAVLFVVFVFM
```

Query Match 53.1% Score 237.4; DB 1; Length 96231;

Best Local Similarity 70.7%; Pred. No. 4.8e-51; Indels 0; Gaps 0;

Matches 316; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
QY 1 ATGTCGATCAGATTATACATCGACTAATTTAGGTGTAAGTCACTGATTAATTTT 60
DB 56074 ATGTTGATCGATTATATATGCAATTAGAGTTACGACTGACCTTATTTT 56015
QY 61 GCGGAAAAACAGGATCGTCCGCGAGACTGTTGTAACCGGATTTAGACTGTG 120
DB 56014 ACAGAAAGACAGATTATCTGAGTTAGTTGTAAGTCACTGATTAATTA 55955
QY 121 TTTATCAGCGGCTTATTTTCTGTTTGTAGAGCTGCTCACTTATGTATC 180
DB 55954 TTTAATCAGCCCGATTTATGTTGTTGTTTATTAATCAATTTTAAATG 55895
QY 181 GTGAATACGTTTATTCGAATTTTGAATTTGTAAGAGCAGCAAAATTCGTC 240
DB 55894 GTTACGTATGTTGTTCAAGATTCATGATTTATATGCGGAGAAATTTG 55835
QY 241 CTGATTAACAGGATCGTCCGTTGATTTTCTATACCGATTTGACATTT 300
DB 55834 CTAAATTAACAGGATTTGTTAAACCTTTATTTGATTTGATTCCTGTT 55775
QY 301 GAAATCGCAGATTTGAGGATTCGATCATCGTCCAGTTTAAATGCAATCA 360
DB 55774 GAGATTTTGAATTCGTTGATTTGAGATTTGTTTCCAGATTAATGCAAT 55715
QY 361 CAGAAACAGGTTTAAACATTAAGTTCCGAGACAGCTGCTATTGAGCGAGCA 420
DB 55714 CAAAGCAAGGTTTACATTAACATTTGGAATCAATTTGTTAAGTGTGCAAC 55655
QY 421 GCTATCATGTTGTTTACTTAAAT 447
DB 55654 GCAATCATGAATTAATTAATTTT 55628
```

Search completed: February 27, 2006, 11:11:04
Job time : 2576.57 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 2522.91 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-2

Sequence score: 447
1 atgttcgacagattata.....tgccttcttactactaact 447

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match. 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_esc1:*

2: gb_esc2:*

3: gb_esc3:*

4: gb_esc4:*

5: gb_esc5:*

6: gb_esc6:*

7: gb_esc7:*

8: gb_esc8:*

9: gb_esc9:*

10: gb_esc10:*

11: gb_esc11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.8	9.6	296	1	AM876109 CML-PT001
2	40.8	9.1	273	10	CL235356 ZMMBB057
3	39.8	8.9	778	10	AG490873 Mus muscu
4	39.6	8.9	258	10	CG907847 ZMMBB052
5	39.6	8.9	905	9	BZ698247 PUBLV12TD
6	39.4	8.8	646	9	BH530334 BOCQZ45TR
7	39.4	8.8	666	10	CL941698 CA ABA005
8	39.4	8.8	771	9	BZ601245 WHACX08TF
9	39.4	8.7	569	9	BH783640 fmbd013f0
10	38.6	8.6	421	1	AUI75308 AUI75308
11	38.6	8.6	568	6	CD323072 STRPUB37.
12	38.4	8.6	764	8	DN041664 JGI CABAI
13	38.4	8.6	887	8	DR732535 FGAS07845
14	38.4	8.6	1007	9	BH135121 ENTNZ59TR
15	38.4	8.5	534	9	AQ765914 HS_5431_A
16	38.4	8.5	779	10	CL150443 104_332_1
17	37.8	8.5	577	9	AZ399037 1M0164N09
18	37.6	8.4	677	7	CK654217 AGENCOURT
19	37.4	8.4	487	6	CF811624 NA838 CDN
20	37.4	8.4	627	10	AG921347 Drosophila
21	37.4	8.4	849	10	DU002865 289737 To
22	37.4	8.4	858	5	BX440759 BX440759

C	23	37.2	8.3	1032	7	CN645534	CN645534 ILUMIGEN
C	24	37	8.3	446	10	CZ073605	CZ073605 OM_Ba006
C	25	37	8.3	490	9	BH267282	BH267282 CH230-138
C	26	37	8.3	694	9	BH355163	BH355163 CH230-81P
C	27	36.8	8.2	397	11	CNS03XCK	AL264845 Tetradon
C	28	36.8	8.2	683	9	BH932950	BH932950 odeb3f12.
C	29	36.6	8.2	1170	11	CNS06QF9	AL410587 T3 end of
C	30	36.6	8.2	641	9	AZ835046	AZ835046 2M0129N03
C	31	36.6	8.2	726	9	AZ317662	AZ317662 1M0036008
C	32	36.6	8.2	808	9	BH096957	BH096957 RPCI-24-2
C	33	36.6	8.2	1712	4	CNS0R4QY	CR655492 Tetradon
C	34	36.4	8.1	361	9	BH756148	BH756148 SALK_0532
C	35	36.2	8.1	618	11	DE054014	DE054014 Oryzias 1
C	36	36.2	8.1	673	1	AV653326	AV653326 AV653326
C	37	36.2	8.1	750	10	CNS011TD	AL100303 Drosophila
C	38	36.2	8.1	961	11	CNS06EOW	AL394518 T7 end of
C	39	36.2	8.1	1030	10	CL056667	CL056667 CH216-83H
C	40	36	8.1	503	9	AQ534563	AQ534563 RPCI-11-3
C	41	36	8.1	694	6	CD904177	CD904177 G356.112K
C	42	36	8.1	738	9	AQ210826	AQ210826 HS_2230_A
C	43	36	8.1	799	5	BX446786	BX446786 BX446786
C	44	36	8.1	932	5	BQ69779	BQ69779 AGENCOURT
C	45	36	8.1	1101	10	CNS0036G	AL064419 Drosophila

ALIGNMENTS

RESULT 1
LOCUS AM876109 296 bp mRNA linear EST 22-MAY-2000
DEFINITION CML-PT0016-231299-075-d08 PT0016 Homo sapiens CDNA, mRNA sequence.
ACCESSION AM876109.1 GI:8014132
VERSION AM876109.1 GI:8014132
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 296)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagal, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V.,
O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
JOURNAL PUBLISHED
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=cl2=CML-PT0016-231
299-075-d08&cl=1999-12-23&cl=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 296.
Location/Qualifiers
1..296
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="PT0016"

FEATURES
Source

Qy 176 TTATGTAATACGCTTATTCATTTATGATTTTGTACGACGAGAAATTCGCTG 235
 Db 567 TTAGCGTGTAAAGGTTTATTCATTTGTTGATTTTCTAA-----AAAAAATATTC 620
 Qy 236 CCATGCTGATACAGGATTCGCTTAAATGCGCTTATTTGATTTATACCGGATG 292
 Db 621 CTCTCTGCTGTTGGGAGATTCATTAAGATGCTCTTTTCTTTTCTTACCTGGGTG 677

RESULT 4

CG907847 259 bp DNA linear GSS 09-DEC-2003
 LOCUS ZMMBB0521G12r ZMMBB (HindIII) Zea mays genomic clone
 DEFINITION ZMMBB0521G12 3', genomic survey sequence.
 CG907847
 ACCESSION CG907847.1 GI:39607114
 VERSION
 KEYWORDS
 SOURCE

ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 258)
 Bhatti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
 Zohovetz,V., Fuks,G., Yu,Y., Wang,R. and Messing,J.
 Sequencing of the maize genome at PGIR (2003c)
 Unpublished (2003)
 CONTACT: Bhatti,A.K.
 DR. Joachim Messing's lab
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
 University
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA
 Tel: 732 445 3801
 Fax: 732 445 5735
 Email: bhatti@waksman.rutgers.edu
 Seq primer: SPE
 Class: BAC ends
 High quality sequence start: 112.
 Location/Qualifiers

FEATURES
 source
 1..258
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultiivar="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBB0521G12"
 /lab_host="E. coli DH10B"
 /clone_lib="ZMMBB (HindIII)"
 /note="Vector: pCUG1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 8.9%; Score 39.6; DB 10; Length 258;
 Best Local Similarity 63.8%; Pred. No. 1.9;
 Matches 60; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 136 TTTATTTTACTGTTTGTGAGAGCTTCTCACTTATGTTATCGTAATACGTTTA 195
 Db 129 TTTGTTTATATGTTTATATAGAAATCATCTAAGATATATCATGAATGAGAGAGA 188
 Qy 196 TCCAAATTTATGATTTTGTACGACGAGCAAAAT 229
 Db 189 AGTTGATTTATGAACCTAGCCCTGAGCAGAAAT 222

RESULT 5
 BZ698247/c 905 bp DNA linear GSS 19-FEB-2003
 LOCUS PUBLV12D ZM 0.6_1.0 KB Zea mays genomic clone ZMMBTA084B24,
 DEFINITION genomic survey sequence.
 BZ698247
 ACCESSION BZ698247.1 GI:26418094
 VERSION
 KEYWORDS
 SOURCE

Zea mays

ORGANISM

Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 905)
 Whitehead,C.A., Quackenbush,J., Van Aken,S., Uteirback,T.,
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
 Bennetzen,J.
 Maize Genomics Consortium
 Unpublished (2003)
 CONTACT: Cathy Whitehead
 TIGR

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 CONTACT: Cathy Whitehead
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitehead@tigr.org
 Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers

FEATURES

source
 1...905
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBTA084B24"
 /clone_lib="ZM 0.6_1.0 KB"
 /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
 cot selected genomic DNA library"

ORIGIN

Query Match 8.9%; Score 39.6; DB 9; Length 905;
 Best Local Similarity 63.8%; Pred. No. 2.3;
 Matches 60; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 136 TTTATTTTACTGTTTGTGAGAGCTTCTCACTTATGTTATCGTAATACGTTTA 195
 Db 770 TTTGTTTATATGTTTATATAGAAATCATCTAAGATATATCATGAATGAGAGAGA 711
 Qy 196 TCCAAATTTATGATTTTGTACGACGAGCAAAAT 229
 Db 710 AGTTGATTTATGAACCTAGCCCTGAGCAGAAAT 677

RESULT 6
 BH530334 646 bp DNA linear GSS 14-DEC-2001
 LOCUS BOGQ245TR BOGQ Brassica oleracea genomic clone BOGQ245, genomic
 survey sequence.

DEFINITION BH530334
 VERSION BH530334.1 GI:17746779
 ACCESSION
 KEYWORDS
 SOURCE

ORGANISM

Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 646)
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
 Uteirback,T.R., Mortman,J.R., White,O.R. and Town,C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)

JOURNAL
 PUBMED
 COMMENT
 Other GSSs: BOGQ245TR
 Contact: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES	Location/Qualifiers
source	1. .646

```

organism="Brassicica oleracea"
mol_type="genomic DNA"
strain="T01000DH3"
db_xref="taxon:3712"
clone="HBOQ245"
clone_1lb="BBOQ"
name="Vector: pHOS1; Site: 2-3 kb sheared
nucleic acid inserted into pHOS1 using BclI linkers

```

ORIGIN

Query Match	8.8%	Score 39.4;	DB 9;	Length 646;
Best Local Similarity	57.9%;	Pred. No. 2.5;		
Matches 70;	Conservative 0;	Mismatches 51;	Indels 0;	Gaps 0;

[illegible]

RESULT 7	
LOCUS	CL941698
DEFINITION	CL941698 666 bp DNA linear GSS 14-SEP-2004
ACCESSION	OA_ABa0054I07.f OA_Aba Oryza australiensis genomic clone OA_ABa0054I07 5', genomic survey sequence.
VERSION	CL941698
KEYWORDS	CL941698.1 GI:52073279
SOURCE	GSS.
ORGANISM	Oryza australiensis Oryza australiensis

FEATURES	Location/Qualifiers
source	1. .666

```

organism="Oryza australiensis"
mol_type="genomic DNA"
db_xref="taxon:4532"
clone="OA_ABA0054107"
tissue_type="young leaves"
lab_host="DH10B T1 phage resistant"
clone_1b="OA_Aaa"
note="vector: pGATAC1; site_1: HindIII; site_2: HindIII"

```

ORIGIN

Query Match 8.8%; Score 39.4; DB 10; Length 666;

Best Local Similarity 55.5%; Pred. No. 2.5;
Matches 76; Conservative 0; Mismatches 61; Indels 0; Gaps 0.

QY 169 ACTATGTTATCGGAATACGGTTTATCCAAATTATGATTGTGACGAGCAGAA 228

Db 491 AATATTTTATGATTAATAAGGTTTTTGAAATATTATCATTTATGATATAGTTAATA 550

QY 229 TTGCGTCGCATGCGATACAGAGGATCGTCCATAAAATCGCGTTGATTTTCTATACCG 288

Db 551 TTCTTTTCCCTCTTAATAGAAATACCTTGGAATAAGATTTTGTATTAATTTTTTCC 610

QY 289 ATGTACCATTTGAAAT 305

Db 611 TTTTATGATTTCAAT 627

RESULT 8	B2601245	B2601245	771 bp	DNA	linear	GSS 08-JUN-2003
LOCUS						
DEFINITION	B2601245	WHAICX08TF Human MCF7 breast cancer cell line library (MCF7_1) Homo sapiens genomic clone MCF7_1-18B16, genomic survey sequence.				
ACCESSION	B2601245	B2601245				
VERSION	B2601245.1	GI:31509707				
KEYWORDS	GSS.					
SOURCE	Homo sapiens	Homo sapiens (human)				
ORGANISM	Homo sapiens					

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 771)	Voilk,S., Zhao,S., Chin,K., Brehner,J.H., Herndon,D.R., Tao,Q., Kombeil,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.	End-sequence profiling: Sequence-based analysis of aberrant genomes	Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)	
		Contact: Voilk SV		
		Colin Collins' lab		
		UCSF Comprehensive Cancer Center		
		UCSF Box 0808, San Francisco, CA 94143-0808, USA		
		Tel: 415 502 7066		
		Fax: 415 502 5665		
		Email: svolk@cc.ucsf.edu		
		This clone is available from Amplicon Express		
		http://www.genomex.com		
		Class: BAC ends.		

```

location/Qualifiers
1..771
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-18B16"
/sex="female"
/clone_lib="human MCF7 breast cancer cell line library
(MCF7_1)"
/notes="Vector: pECBAC1, site_1: HindIII; This library was
constructed from MCF7 breast cancer cell line by AmpliCon
Express (http://www.genomex.com) using their standard
procedure."
```

ORIGIN

Query Match	8.8%	Score 39.4;	DB 9;	Length 771;
Best Local Similarity	47.1%;	Pred. No. 2.6;		
Matches 121; Conservative	0;	Mismatches 136;	Indels 0;	Gaps 0

OY	80	TGCGGAGAGACTGGTTGTACCGGGATTTTAAAGCATGTCGTTAAATCAGCGCGATCTTAA	139
Db	75	TGTATCTAGCCTCATTTATATATATAAAGATCAGCATCTTTTTTAAATAAAGCTTTTTT	134
OY	140	TTTTACTTGTTTTTCCTAGTAGACTTCTCTCACTTATGTTATGTCGAATATAGGTTTATTCGA	199
Db	135	GTCAGTTCCTATTATGGCTATTTTGGAAATTTATATTTTGTGAGATACAGTATTTGTTGA	194

QY 200 AATTATGATTTTGTACGACGACAGAAATTCGTCGATGCTATAAGGAGATGTC 259
 DB 195 GTTTTCCTTTGATTTCTAAATAACACTTTCCTTTCTATTATGTAAGATTATGTC 254
 QY 260 TAAATATGCGCTGATTTTCTATACCCGATTTGACATTTGAATTCGAGATTTGAG 319
 DB 255 TTATATTCACATTTTAATTCCTTTGGAGATTTTAAGATTTATTCACAAACTACTT 314
 QY 320 GAATCGCATCATCGTG 336
 DB 315 TTCTCTCATATTGTG 331

RESULT 9

BH783640 569 bp DNA linear GSS 28-MAR-2002
 LOCUS fzm013f008g12k0 fzm filtered library Zea mays genomic clone
 DEFINITION fzm013f008g12 5', genomic survey sequence.

ACCESSION BH783640 GI:19787191
 VERSION BH783640.1
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACMAD
 clade; Panicoideae; Andropogonae; Zea.
 AUTHORS 1 (bases 1 to 569)
 TITLE Buddman, M.A., Freese, R.G., Bedell, J.A., Nundberg, A.N. and Lakey, N.D.
 JOURNAL Genethresher methylation filtered genomic sequences from maize
 COMMENT Unpublished (2002)
 CONTACT: Bedell JA
 ORION Genomics, LLC
 4041 Forest Park Ave, St. Louis, MO 63108, USA
 Tel: 314 615 6979
 Fax: 314 615 5975
 Email: jbedell@oriongenomics.com
 Plate: fzm013f008 row: 9 column: 12
 Seq primer: SK reverse
 Class: methylation filtered
 High quality sequence stop: 569.
 Location/Qualifiers

FEATURES
 source

1..569
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="inbred line Mo17"
 /db_xref="taxon:4577"
 /clone="fzm013f008g12"
 /clone_lib="fzm filtered library"
 /note="Organ: leaf; Vector: pBSK(-); Site 1: HincII; DNA
 prepared from purified nuclei was randomly sheared,
 end-repaired, size fractionated to enrich for the 0.5 to
 5 kb fraction, ligated into HincII-digested pBSK(-)
 vector and electroporated into E. coli cells."

ORIGIN

Query Match 8.7%; Score 39; DB 9; Length 569;
 Best Local Similarity 51.4%; Pred. No. 3.2;
 Matches 90; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 40 GATCTACTAGTTTAATTTTGGCGAAAAACAGGATCGTGGCGGACAGACTTTGTA 99
 DB 272 GAATCAACAGGTTTACCTGTAATAGAAAGCATAGAGAGCCCTGATAGATCTTGTGT 213
 QY 100 CCGGATATTTAGACTGTGTTTAATACACCGGCTTTATTTACTGTTTGTCTAGTG 159
 DB 212 ACTGATTTCTAATGAGAGGTTAAGACATTTTGACCTTACAGGTTTGCTACTT 153
 QY 160 AGCTGCTCACTTATGTTATCGTAATACGTTTATCCAATTTATGATTTGT 214
 DB 152 TGGCTTCAGCTTTTGTGTGCTCCAGCTTAGCTGAATGTCCTTTT 98

RESULT 10

AUI75308 421 bp mRNA linear EST 11-JUN-2003
 LOCUS AUI75308
 DEFINITION AUI75308 Marsupenaeus japonicus adult Marsupenaeus japonicus cDNA
 clone Pui366 forward similar to Leshmania major Lmlv393/298A
 Leshmania major promastigote full length cDNA library from early,
 mRNA sequence.

ACCESSION AUI75308
 VERSION AUI75308.1 GI:14621715
 KEYWORDS EST.
 SOURCE Marsupenaeus japonicus
 ORGANISM Marsupenaeus japonicus

REFERENCE

AUTHORS Rojlimakorn, J., Hirano, I., Itami, T., Takahashi, Y. and Aoki, T.
 TITLE Gene expression in haemocytes of Kuruma prawn, Penaeus japonicus,
 in response to infection with WSSV by EST approach
 JOURNAL Fish Shellfish Immunol. 13 (1), 69-83 (2002)
 PUBMED 12201653

COMMENT

Contact: Takashi Aoki
 Aquatic Biosciences
 Tokyo University of Fisheries
 Komae 4-5-7, Tokyo, Minato-ku 108-8477, Japan
 Tel.: 81-3-5463-0689
 Fax: 81-3-5463-0690
 Email: aoki@tokyo-u-fish.ac.jp.
 Location/Qualifiers

FEATURES

source
 1..421
 /organism="Marsupenaeus japonicus"
 /mol_type="mRNA"
 /db_xref="taxon:27405"
 /clone="Pui366"
 /cell_type="hemocytes"
 /dev_stage="adult"
 /clone_lib="Marsupenaeus japonicus adult"
 /note="Vector: lambda Zap II"

ORIGIN

Query Match 8.6%; Score 38.6; DB 1; Length 421;
 Best Local Similarity 36.7%; Pred. No. 4;
 Matches 81; Conservative 35; Mismatches 105; Indels 0; Gaps 0;

QY 134 TCTTATTTTACTGTTTGTGTCAGTGCCTGCTACTATGTATGCGAATACGGTT 193
 DB 1 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 60
 QY 194 TATCCAAATTTATGATTTTGTACGACGACGAAATTCGCTGCATGCTGATTAACAGGA 253
 DB 61 YYYCAATTAATAAKTWTGCAAGGCMCTTMMAGCASCMTGTATTAATTTKAGYTTKKYCY 120
 QY 254 TCGTCTTAATAATTCGGTTGATTTTCTATACCCGATTTGACATTTGAATTCGCAAT 313
 DB 121 YCYCCAKTKGSGGSYTTGGMWTTMMWTTTWTWTTTTCATTAASGKTYCYAHMYCC 180
 QY 314 TTGAGAGAAATCGCATCATCGTCCAGGTTTAATTCGCAT 354
 DB 181 WTGKGGAAMAGSCYTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 221

RESULT 11

CD323072 566 bp mRNA linear EST 17-SEP-2003
 LOCUS CD323072
 DEFINITION Strp537.001434 Sea urchin embryo 20th blastula stage cDNA library
 MPMG537 Strongylocentrotus purpuratus cDNA clone
 CALDB53701524; MPl_537_24015 5', mRNA sequence.

ACCESSION CD323072 GI:34795133
 VERSION CD323072.1
 KEYWORDS EST.
 SOURCE Strongylocentrotus purpuratus
 ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinoidae; Euechinoidae; Echinacea; Echinoida;

REFERENCE Strongylocentrotidae: Strongylocentrotus.
1 (bases 1 to 568)
AUTHORS Pousetka,A.J., Groth,D., Hennig,S., Thamm,S., Cameron,A., Beck,A., Reinhardt,R., Herwig,R., Panopoulou,G. and Lehnach,H.
TITLE Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters
JOURNAL Genome Res. 13 (12), 2736-2746 (2003)
COMMENT Contact: Pousetka AJ
Laboraty 145, dept. Lehnach
Max-Planck-Institut fuer Molekulare Genetik
Inhestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: pousetka@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONP) to reduce sequencing redundancy. According to the ONP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONP cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the other clones assigned to the same ONP cluster as the clone from which the above EST is generated is available at the sea urchin project web site at: http://www.molgen.mpg.de/ag_seaurchin/. cDNA clones and filters are distributed via the Resource Center/Primary Database of the German Human Genome Project (<http://www.rzpd.de>)
PCR primers
FORWARD: 5' CCCAGGCTTACACTTATGCTTCCGGCTG 3' (M13SP) 5'-seq
BACKWARD: 5' GCTATACGCCAGTCGCGAAGGGGATGTG 3' (M13SP) 3'-seq
Seq primer: 5'-CCGATCCGGAAATCCCGGT-3' pSPORT3/86
High quality sequence stop: 568.
Location/Qualifiers
1..568
/organism="Strongylocentrotus purpuratus"
/mol_type="mRNA"
/db_xref="taxon:7668"
/clone="CALTP53701524;MPI 537_24015"
/tissue_type="whole embryo"
/dev_stage="embryonic 20hr"
/lab_host="E.coli, XLI Blue"
/clone_lib="Sea urchin embryo 20hr blastula stage cDNA library MPMP537"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; Random primed and directionally cloned in pSPORT1 vector using a NotI (5'-pCACTAGTCTAGATGCGACGCGCCG (T)15-3' and a SalI 5'-TCGACCCACGCGTCCG-3' adaptors (Gibco BRL) "
ORIGIN
Query Match 8.6%; Score 38.6; DB 6; Length 568;
Best Local Similarity 51.4%; Pred. No. 4.2; Mismatches 84; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 133 GCTTTATTTTACTGTTTGTCTAGTGAAGCTTGTCACTATGTTATGTTGAATACGCT 192
DB 8 GTCCGTGCTTATGTTGTCTTAAGTCTCTGTTTAACATGTTCTGTTCATCTGT 67
QY 193 TTATCCAAATTATATATTTGTACGACGCCAGAAAATTCGCTGCATGCTGATAACAGG 252
DB 68 AGATTATATCATCATATATATATATATTCAGATTAATTTGAATATCTTCCTTATATCCTTAGG 127
QY 253 ATCGTCTAAATAATGCGTTGATTTTCTATACCCGATTTGATACATTTGCAAT 305
DB 128 TTGTATATATATTTCAATTTTCTTTTGTGTTGCTGATTTGCTTTTATATAT 180

RESULT 12
DN041664 764 bp mRNA linear EST 10-FEB-2005
LOCUS DN041664/c
DEFINITION JGI_CABAI1336.rev NIH_XGC_tropikid1 xenopus tropicalis cDNA clone
IMAGE:7746695 3', mRNA sequence.
ACCESSION DN041664
VERSION DN041664.1 GI:59247540
KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 764)
AUTHORS Richardson,P., Lucas,S., Rokhsar,D., Dettler,J.C., Ng,D.C., Brokstein,P. and Lindquist,E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNAL Unpublished (2004)
COMMENT Other_ESTs: JGI_CABAI136.fwd
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Bruce Blumberg Laboratory, University of California, Irvine
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNL:
<http://image.llnl.gov>
Naming Conventions: EST name is generated by the concatenation of the JGI Clone id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
Plate: CAB 0013 row: p column: 21
High quality sequence stop: 756.
Location/Qualifiers
1..764
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/strain="N6 (Nigerian 6th generation inbred)"
/db_xref="taxon:8364"
/clone="IMAGE:7746695"
/tissue_type="Kidney"
/dev_stage="Adult"
/lab_host="Electromax DH10B T1 Phage Resistant cells"
/clone_lib="NIH_XGC tropikid1"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dT priming
(5'-GACAGAGAGAGAGAGAGACTGTCGAGTTTGTGTTTGTGCTAGTGAAGT 3') and StrataScript reverse transcriptase. After ligation of EcoRI adaptors (5'-AATTCGACAGACG-3') followed by kinasg adaptors and by XhoI digestion, the cDNA was size selected by chromatography on Sepharose CL-2B columns and fractions containing cDNAs larger than 1000 bp were ligated into EcoRI/XhoI-digested pCS107. Reference for library construction: Current Genomics 4, 635-644. Library constructed by Michelle Tabb and Bruce Blumberg (Dept of Developmental and Cell Biology, University of California, Irvine). "
ORIGIN
Query Match 8.6%; Score 38.4; DB 8; Length 764;
Best Local Similarity 58.9%; Pred. No. 5; Mismatches 46; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 104 GATATTTAGACTGTGTGTTTATATGACCGGCTTTATTTTACTGTTTGTCTAGTGAAGT 163
DB 400 GAGTTTATGTTTCTGTATATCCAGTCAGATATATATATATTTGTGTGAGTGCCT 341
QY 164 TGGTCACTTATGTTATGCGTAATACGTTTATCCAAATTTATGATTTTGA 215
DB 340 CTCCTATATTATGTTACAGATATATATATATACAGATATATATATATA 289

RESULT 13
DR732535 887 bp mRNA linear EST 18-JUL-2005
LOCUS DR732535
DEFINITION FGAS078455 Triticum aestivum FGAS: Tates Triticum aestivum cDNA,

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time 282.194 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatcagattata.....tgttcgttactactaatt 447

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2002as:*
6: geneseqn2003as:*
7: geneseqn2004as:*
8: geneseqn2005as:*
9: geneseqn2006as:*
10: geneseqn2007as:*
11: geneseqn2008as:*
12: geneseqn2009as:*
13: geneseqn2010as:*
14: geneseqn2011as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	447	100.0	447	10 ADA44792	ADA44792 Bacillus
2	447	100.0	447	10 ADA007642	ADA007642 B subtilis
3	447	100.0	447	13 ADV65733	ADV65733 B. subtilis
4	447	100.0	6536	10 ADA44796	ADA44796 Microbial
5	440.6	98.6	450	4 AAF82256	AAF82256 Bacillus
6	440.6	98.6	3045	4 AAF82254	AAF82254 Bacillus
7	289.2	64.7	450	6 ABK75634	ABK75634 Bacillus
8	189.4	42.4	453	4 AAH53465	AAH53465 S. epider
9	189.4	42.4	459	6 ABN92203	ABN92203 Staphyloc
10	189.4	42.4	459	13 ADS01974	ADS01974 Staphyloc
11	182	40.7	3240	4 AAH53379	AAH53379 S. epider
12	65.2	14.6	480	9 ADB06111	ADB06111 Altiolococ
13	65.2	14.6	480	9 ADB06109	ADB06109 Altiolococ
14	65.2	11.8	110000	9 ADB12064	ADB12064 Altiolococ
15	52.8	10.2	198	6 ABN92026	ABN92026 Staphyloc
16	45.6	9.7	2000	11 AAH54833	AAH54833 S. epider
17	43.2	8.4	2000	11 ACI37108	ACI37108 Rice stre
18	37.4	8.1	5822	6 ABL33096	ABL33096 Human imm
19	36.2	8.1	26997	4 AAS46748	AAS46748 Tumour su

20	36.2	8.1	110000	10 ADF77343_06	Continuation (7 of
21	36	8.1	7195	4 AAS45325	AAS45325 Chemical1
22	36	8.1	7195	6 ABR28166	ABR28166 DNA trans
23	35.2	7.9	2100	6 ABR72930	ABR72930 Bacillus
24	35.2	7.9	4078	6 AAD29896	AAD29896 B. lichen
25	35	7.8	1347	8 ACA49970	ACA49970 Prokaryot
26	35	7.8	2116	12 ADN12148	ADN12148 Interleuk
27	35	7.8	2116	13 ACF87528	ACF87528 Human SIR
28	35	7.8	2116	14 AD226514	AD226514 Human IL7
29	35	7.8	11839	4 AAL04940	AAL04940 Human rep
30	35	7.8	11839	4 ABL97834	ABL97834 Human tes
31	35	7.8	11839	8 ABR27494	ABR27494 Secreted
32	35	7.8	11839	8 ADA44503	ADA44503 Human sec
33	34.8	7.8	2000	11 ACL37886	ACL37886 Rice stre
34	34.8	7.8	2000	11 ACL35887	ACL35887 Rice stre
35	34.4	7.7	534	6 AAF24848	AAF24848 Degenerat
36	34.4	7.7	534	6 AAD42512	AAD42512 Human zel
37	34.4	7.7	534	9 ACH00107	ACH00107 Dehgenera
38	34.4	7.7	5769	6 ABR40019	ABR40019 Human che
39	34.4	7.7	51259	2 AAX83007	AAX83007 Partial m
40	34.2	7.7	1344	10 ABR07028	ABR07028 S. pneumo
41	34.2	7.7	5273	6 ABL32874	ABL32874 Human imm
42	34.2	7.7	6052	6 ABL32417	ABL32417 Human imm
43	34.2	7.7	12666	2 AAV52270	AAV52270 Streptoco
44	34.2	7.7	110000	10 ABS56454_11	Continuation (12 o
45	34.2	7.7	277616	13 ABD32602	ABD32602 Human can

ALIGNMENTS

RESULT 1	ADA44792	ADA44792 standard; DNA; 447 BP.
ID	ADA44792	ADA44792 standard; DNA; 447 BP.
XX	ADA44792;	
DT	20-NOV-2003	(first entry)
XX	Bacillus subtilis poly-gamma-glutamate synthetase complex pgsC gene.	
XX	Cell surface; expression vector; microbial;	
KW	poly-gamma-glutamate synthetase; pgsBCA complex; surface expression;	
KW	Gram-positive bacterium; Gram-negative bacterium; enzyme; antigen;	
KW	antibody; attachment protein; adsorption protein; vaccine; pgsC; gene;	
OS	Bacillus subtilis; variety chungkookjang.	
PN	WO2003014360-A1.	
PN	20-FEB-2003.	
PD	09-AUG-2002; 2002WO-KR001522.	
PF	10-AUG-2001; 2001KR-00048373.	
PR	(BIOL-) BIOLEADERS CORP.	
PA	(MDMD-) MD LAB CO LTD.	
PI	Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashiuchi M;	
PI	WPI; 2003-256589/25.	
PT	New expression vector containing gene(s) that encode a poly-gamma-	
PT	glutamate synthetase complex, useful for producing proteins (e.g.	
PT	vaccines or enzymes) on the microbial surface of Gram-positive and/or	
PT	Gram-negative bacteria.	
PS	Claim 3; Page 111; 122pp; English.	
XX	The invention relates to a vector for expression of a target protein on a	
CC	microbial cell surface. The vector of the invention comprises either one	

or more than two genes of the *Bacillus subtilis* poly-gamma-glutamate synthetase complex (pgsBCA) to facilitate microbial surface expression of the target protein. The pgsBCA gene complex comprises the pgsB, pgsC and pgsA genes and is normally expressed in the outer membrane of *Bacillus subtilis*. The vector can be transformed into either Gram-positive or Gram-negative bacteria (e.g., *Escherichia coli*), and can be used for the surface expression of various proteins of interest such as enzymes, antigens, antibodies, attachment proteins or adsorption proteins. Proteins recombinantly produced using the vector of the invention can be used as, for example, vaccines or enzymes. The present sequence represents the *Bacillus subtilis* pgsBCA complex gene pgsC, which is specifically claimed for use in the vector of the invention.

Sequence 447 BP; 116 A; 84 C; 94 G; 153 T; 0 U; 0 Other;

Query Match 100.0%; Score 447; DB 10; Length 447;

Best Local Similarity 100.0%; Pred. No. 2.7e-126;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ATGTTCCGATCAGATTATATACATGCACTAATTTAGGAGTACTGAGTTTAAATTTT 60
DB 1 ATGTTCCGATCAGATTATATACATGCACTAATTTAGGAGTACTGAGTTTAAATTTT 60
OY 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
DB 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
OY 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
DB 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
OY 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
DB 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
OY 181 GTGAATACGAGTTTATCCAAATTTATGATTTGTACGAGACGAGAAATTTGGCTGCATG 240
DB 181 GTGAATACGAGTTTATCCAAATTTATGATTTGTACGAGACGAGAAATTTGGCTGCATG 240
OY 241 CTGATTAACAGGATCGTCTTAAATTCGCTTTGATTTTCTATACCCGATTTGACATTT 300
DB 241 CTGATTAACAGGATCGTCTTAAATTCGCTTTGATTTTCTATACCCGATTTGACATTT 300
OY 301 GAAATCGAGAAATTCGAGGAATCGGCATCATGTCGACAGTTTAAATTCGCAATTCATTT 360
DB 301 GAAATCGAGAAATTCGAGGAATCGGCATCATGTCGACAGTTTAAATTCGCAATTCATTT 360
OY 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGACAGCTGCTATTGAGGAGGAGCACTTT 420
DB 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGACAGCTGCTATTGAGGAGGAGCACTTT 420
OY 421 GCTATCATGTTTGTACTACTTAATT 447
DB 421 GCTATCATGTTTGTACTACTTAATT 447
```

RESULT 2

ADO07642 ID ADO07642 standard; DNA; 447 BP.

XX ADO07642;

DT 15-JUL-2004 (first entry)

DE B subtilis poly-X-glutamate synthetase complex coding sequence pgsC.

XX ds; gene; enzyme; vaccine; cytosolic; pgsC; poly-X-glutamate synthetase;

KM human papilloma virus.

XX *Bacillus subtilis*.

OS WO0004035795-A1.

XX 29-APR-2004.

PF 17-OCT-2003; 2003WO-KR002163.

XX

PR 17-OCT-2002; 2002KR-00063378.

XX

PA (BIOL-) BIOLEADERS CORP.

PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

XX

PI Sung M, Poo H, Lee J, Jung C, Hong S, Kim C, Park S, Pyo H;

XX WPI; 2004-348463/32.

XX

PT New vector containing pgs A-C genes encoding poly-gamma-glutamate

PT synthetase complex and an antigen protein gene of human papilloma virus,

PT useful in preparing vaccine for treating or preventing mucosal tumor,

PT e.g. cervical cancer.

XX

PS Disclosure; Page 61-62; 69pp; English.

XX

CC The present invention relates to a vector for preparing a vaccine which

CC contains one or more than two genes, i.e. pgs A-C encoding poly-X-

CC glutamate synthetase complex and an antigen protein gene of human

CC papilloma virus. The vector and microbes transformed with it are useful

CC in preparing vaccines for treating or preventing mucosal tumor, e.g.

CC cervical cancer. The present sequence is a *Bacillus subtilis* poly-X-

CC glutamate synthetase complex coding sequence.

XX

Sequence 447 BP; 116 A; 84 C; 94 G; 153 T; 0 U; 0 Other;

Query Match 100.0%; Score 447; DB 12; Length 447;

Best Local Similarity 100.0%; Pred. No. 2.7e-126;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ATGTTCCGATCAGATTATATACATGCACTAATTTAGGAGTACTGAGTTTAAATTTT 60
DB 1 ATGTTCCGATCAGATTATATACATGCACTAATTTAGGAGTACTGAGTTTAAATTTT 60
OY 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
DB 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
OY 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
DB 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
OY 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
DB 121 TTTATATACGCGGCTTTATTTTATCTGTTTGTAGAGAGCTTCTCACTTATGTC 180
OY 181 GTGAATACGAGTTTATCCAAATTTATGATTTGTACGAGACGAGAAATTTGGCTGCATG 240
DB 181 GTGAATACGAGTTTATCCAAATTTATGATTTGTACGAGACGAGAAATTTGGCTGCATG 240
OY 241 CTGATTAACAGGATCGTCTTAAATTCGCTTTGATTTTCTATACCCGATTTGACATTT 300
DB 241 CTGATTAACAGGATCGTCTTAAATTCGCTTTGATTTTCTATACCCGATTTGACATTT 300
OY 301 GAAATCGAGAAATTCGAGGAATCGGCATCATGTCGACAGTTTAAATTCGCAATTCATTT 360
DB 301 GAAATCGAGAAATTCGAGGAATCGGCATCATGTCGACAGTTTAAATTCGCAATTCATTT 360
OY 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGACAGCTGCTATTGAGGAGGAGCACTTT 420
DB 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGACAGCTGCTATTGAGGAGGAGCACTTT 420
OY 421 GCTATCATGTTTGTACTACTTAATT 447
DB 421 GCTATCATGTTTGTACTACTTAATT 447
```

RESULT 3

ADV65733 ID ADV65733 standard; DNA; 447 BP.

XX ADV65733;

XX 10-FEB-2005 (first entry)

DT B. subtilis poly-gamma-glutamic acid synthase C gene SEQ ID NO:2.

XX

KW ds; poly-gamma-glutamic acid synthase C; pgsC; gene expression;
 KW antibiotic; antimicrobial; fungicide; cyclostatic.
 OS Bacillus subtilis.
 PN KR2004034780-A.
 XX
 XX 29-APR-2004.
 PD 17-OCT-2002; 2002KR-00063379.
 XX PF 17-OCT-2002; 2002KR-00063379.
 XX PR 17-OCT-2002; 2002KR-00063379.
 XX PA (BIOL-) BIOLEADERS CORP.
 XX PA (UYCH-) UNIV CHUNGSUN CO LTD.
 XX PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
 XX
 XX BoO HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
 PI Park YG, Sung MH;
 DR WPI; 2004-577380/56.
 XX
 PT Method for surface expression of peptides p5 and anal3 using pgs bca
 PT gene.
 XX
 XX PS Disclosure; SEQ ID NO 2; 25pp; Korean.
 CC The invention relates to a novel method for surface expression of
 CC peptides p5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs)BCA
 CC gene, thereby removing a purification process of peptides p5 and Anal3,
 CC and using lactic acid bacteria for the surface expression, so that
 CC peptide antibiotics can be cheaply and stably mass-produced. An
 CC expression vector pHEC1LB:pgsA-p5 comprises one or more genes encoding
 CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
 CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
 CC and anticancer activities, wherein the dipolar peptide antibiotic has
 CC homology to the peptide p5 encoded by the nucleotide sequence set forth
 CC in ADV65735, or to the peptide Anal3 encoded by the nucleotide sequence
 CC set forth in ADV65737. The present sequence represents the B. subtilis
 CC poly-gamma-glutamic acid synthase C gene used in the invention.
 XX
 XX Sequence 447 BP; 116 A; 84 C; 94 G; 153 T; 0 U; 0 Other;
 SQ
 Query Match 100.0%; Score 447; DB 13; Length 447;
 Best Local Similarity 100.0%; Pred. No. 2.7e-126;
 Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGGATCAGATTATATCATCGCATTAATTTAGTGTACTACTCAGTTAATTTT 60
 DB 1 ATGTTGGATCAGATTATATCATCGCATTAATTTAGTGTACTACTCAGTTAATTTT 60
 QY 61 GCGGAAAAACAGGATCGCGCGGCGGAGCTTGTGTACCGGATTTAGGACTTGTG 120
 DB 61 GCGGAAAAACAGGATCGCGCGGCGGAGCTTGTGTACCGGATTTAGGACTTGTG 120
 QY 121 TTTAATCAGCGGCTTTATTTTACTGTTTGTAGTGAAGCTTGTCTCATTTATGTTATC 180
 DB 121 TTTAATCAGCGGCTTTATTTTACTGTTTGTAGTGAAGCTTGTCTCATTTATGTTATC 180
 QY 181 GTGAATACGGATTATTCGAAATTTGATGTTGTAGGAGCGGAAATTCGCTGCATG 240
 DB 181 GTGAATACGGATTATTCGAAATTTGATGTTGTAGGAGCGGAAATTCGCTGCATG 240
 QY 241 CTGATATACAGGATCGCTTAAATGCGGTTGATTTCTAATACCGGATTTGACATTT 300
 DB 241 CTGATATACAGGATCGCTTAAATGCGGTTGATTTCTAATACCGGATTTGACATTT 300
 QY 301 GAAATCGCAGATTTTCAGGATTCGAGATCATCGTGCAGGTTTAATTCGCAATACATT 360
 DB 301 GAAATCGCAGATTTTCAGGATTCGAGATCATCGTGCAGGTTTAATTCGCAATACATT 360
 QY 361 CAGAAACAAGTTTAACATTAGCTTGGAGACGCGTGTCTATTAGCGGAGCACTTT 420
 DB 361 CAGAAACAAGTTTAACATTAGCTTGGAGACGCGTGTCTATTAGCGGAGCACTTT 420

DB 361 CAGAAACAAGTTTAACATTAGCTTGGAGACGCGTGTCTATTAGCGGAGCACTTT 420
 QY 421 GCTATCAGCTTGTGTACTACTAATT 447
 DB 421 GCTATCAGCTTGTGTACTACTAATT 447
 XX
 XX
 XX RESULT 4
 XX ADA44796
 XX ID ADA44796 standard; DNA; 6536 BP.
 XX
 XX AC ADA44796;
 XX
 XX DT 20-NOV-2003 (first entry)
 XX
 XX DE Microbial cell surface expression vector pGNBCA, SEQ ID NO:6.
 XX
 XX KW Cell surface; expression vector; microbial;
 KW poly-gamma-glutamate synthetase; Bacillus subtilis pgsBCA complex;
 KW surface expression; Gram-positive bacterium; Gram-negative bacterium;
 KW enzyme; antigen; antibody; attachment protein; adsorption protein;
 KW vaccine; pGNBCA; cyclic; circular; ds.
 XX
 XX OS Synthetic.
 XX Bacillus subtilis; variety chungkookjang.
 XX
 XX PN WO2003014360-A1.
 XX
 XX PD 20-FEB-2003.
 XX
 XX PF 09-AUG-2002; 2002W0-KR001522.
 XX
 XX PR 10-AUG-2001; 2001KR-00048373.
 XX
 XX PA (BIOL-) BIOLEADERS CORP.
 XX PA (MDMD-) MD LAB CO LTD.
 XX
 XX PI Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashiuchi M;
 XX
 XX DR WPI; 2003-256589/25.
 XX
 XX PT New expression vector containing gene(s) that encode a poly-gamma-
 PT glutamate synthetase complex, useful for producing proteins (e.g.,
 PT vaccines or enzymes) on the microbial surface of Gram-positive and/or
 PT Gram-negative bacteria.
 XX
 XX PS Example 1; Page 113-120; 122pp; English.
 XX
 XX CC The invention relates to a vector for expression of a target protein on a
 CC microbial cell surface. The vector of the invention comprises either one
 CC or more than two genes of the Bacillus subtilis poly-gamma-glutamate
 CC synthetase complex (pgsBCA) to facilitate microbial surface expression of
 CC the target protein. The pgsBCA gene complex comprises the pgsB, pgsB and
 CC pgsA genes and is normally expressed in the outer membrane of Bacillus
 CC subtilis. The vector can be transformed into either Gram-positive or Gram
 CC negative bacteria (e.g., Escherichia coli), and can be used for the
 CC surface expression of various proteins of interest such as enzymes,
 CC antigens, antibodies, attachment proteins or adsorption proteins.
 CC Proteins recombinantly produced using the vector of the invention can be
 CC used as, for example, vaccines or enzymes. The present sequence
 CC represents a microbial cell surface expression vector of the invention,
 CC pGNBCA.
 XX
 XX SQ Sequence 6536 BP; 1855 A; 1464 C; 1567 G; 1649 T; 0 U; 1 Other;
 SQ
 Query Match 100.0%; Score 447; DB 10; Length 6536;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGGATCAGATTATATCATCGCATTAATTTAGTGTACTACTCAGTTAATTTT 60
 DB 1423 ATGTTGGATCAGATTATATCATCGCATTAATTTAGTGTACTACTCAGTTAATTTT 1482

XX New nucleic acid encoding a glutamate racemase useful for the
PT preparation of poly-gamma-glutamic acid.
XX
PS Claim 1; Page 7-11; 17pp. Japanese.

CC The present sequence encodes a glutamate racemase enzyme which is useful
CC in the production of poly-gamma-D-glutamic acid (L-PCA) or poly-gamma-D,
CC glutamic acid (PCA). A plasmid comprising the present sequence may be
CC used to transform *Escherichia coli*. The transformants express the enzyme
CC and PCA is produced in the culture
XX
XX Sequence 3045 BP; 1000 A; 593 C; 636 G; 756 T; 0 U; 0 Other;

SQ Sequence 3045 BP; 1000 A; 593 C; 696 G; 756 T; 0 U; 0 Other;

Query Match	98.6%	Score	440.6	DB	4	Length	3045
Best Local Similarity	99.1%	Pred.	No. 4.9e-124				
Matches	443	Conservative	0	Mismatches	4	Indels	0
						Gaps	0

Oy	1	ATGTCGGATCAGATTATTTATCATCGCACTAATTTTAGGTACAGTACAGTTTAATTTT	60
Db	1215	ATGTCGGATCAGATTATTTATCATCGCACTAATTTTAGGTACAGTACAGTTTAATTTT	1274
Oy	61	GGGAAAAAACAGGAGATGTCGCCGACAGACTTGTGTACCGGGATATTAGACCTTGTG	120
Db	1275	GGGAAAAAACAGGAGATGTCGCCGACAGACTTGTGTACCGGGATATTAGACCTTGTG	1334
Oy	121	TTTATTCAGCCGGCTTTATTTTACTGTGTTTGTGAGTGAAGCTTGCTCACTAATGATATC	180
Db	1335	TTTATTCAGCCGGCTTTATTTTACTGTGTTTGTGAGTGAAGCTTGCTCACTAATGATATC	1394
Oy	181	GTGAATAACGGTTTATCCAAATTTTATGATTTTGTACGAGACGAGAAAATTGCTGCCATG	240
Db	1395	GTGAATAACGGTTTATCCAAATTTTATGATTTTGTACGAGACGAGAAAATTGCTGCCATG	1454
Oy	241	CTGATTAACAGGGAATGCTCTTAAAAATTCGGCTTGAATTTTCTATACCAGATTGACATTT	300
Db	1455	CTGATTAACAGGGAATGCTCTTAAAAATTCGGCTTGAATTTTCTATACCAGATTGACATTT	1514
Oy	301	GAATTCGCGAATTTTCAGAGAAATCGGCAATCATCGAGCCAGGTTTATTTGCCAATACCAAT	360
Db	1515	GAATTCGCGAATTTTCAGAGAAATCGGCAATCATCGAGCCAGGTTTATTTGCCAATACCAAT	1574
Oy	361	CAGAAACAAGTTTAAACCATTAAGTTCCGAGACAGCTGCTATTTGAGCGGAGCGACCTTT	420
Db	1575	CAGAAACAAGTTTAAACCATTAAGTTCCGAGACAGCTGCTATTTGAGCGGAGCGACCTTT	1634
Oy	421	GCTATCATGTTGTTTACTACTAATTT	447
Db	1635	GCTATCATGTTGTTTACTACTAATTT	1661

RESULT 7
ABK75634
ID ABK75634 standard; DNA; 450 BP

AC ABK75634;

DT 13-AUG-2002 (first entry)

DE *Bacillus licheniformis* genomic sequence tag (GST) #2925

KW . Differential gene expression; genomic sequenced tag; GST
KW altered culture condition; environmental stress;
KW physiological provocation; ds.

OS *Bacillus licheniformis*.

PN WO200229113-A2

PD 11-APR-2002

PF 05-OCT-2001; 2001WO-US031437.

XX. (NOVO) NOVOZYMES BIOTECH INC
PA (NOVO) NOVOZYMES AS.
PA Berka R, Clausen IG;
XX
XX
PI
XX
XX
XX MPI, 2002-416684/44.

PI	Berka R, Clausen IG,
XX	
DR	WPI; 2002-416684/44.
XX	

Monitoring differential expression of several genes in first *Bacillus*
cell relative to expression of same genes in one or more second *Bacillus*
cells, by using substrate containing *Bacillus* genomic sequenced tag
array.

PS Claim 4; SEQ ID NO 2925; 200pp; English

The invention describes a method for monitoring differential expression of genes in a first *Bacillus* cell relative to expression of the genes in other *Bacillus* cells, comprising hybridising labelled nucleic acid probes isolated from *Bacillus* cells to a substrate containing array of *Bacillus* genomic sequenced tags (GST), examining the array, and determining relative gene expression by an observed hybridisation reporter signal of a spot in the array. The method is useful for measuring the expression of genes in a first *Bacillus* cell relative to expression of the same genes in one or more second *Bacillus* cells. The method is useful for monitoring global expression of several genes from a *Bacillus* cell, discovering new genes, identifying possible functions of unknown open reading frames and monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which *Bacillus* cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at http://wipo.int/pub/published_pct_sequences

Sequence 450 BP; 112 A; 85 C; 102 G; 151 T; 0 U; 0 Other;

Query Match	64.7%	Score 289.2;	DB 6	Length 450;
Best Local Similarity	78.0%	Pred. No. 5,1e-78;		
Matches 348;	Conservative 0;	Mismatches 98;	Indels 0;	Gaps 0;

QY 1 ATGTTGGATCAGATTATTTAATACATGCACCTAAATTTTAAAGGATGACACTCAGTTAATTTT 60
 Db 1 AAGTTTGGATCAGATTATTAATATACGGCCCTCATTTTAAAGAGTCTTACTCAGTTTGAATTTT 60
 QY 61 GCGGAAAAAACAAGGATCGTGC CGGACAGGACTTGTGTATCCGGGAAATTTAAGACTTGTG 120
 Db 61 GAGAGAAAAACGGGAATGTATCAGCCGGCCCTGTCGTATACGGGTTAATTTGGAGACTTGTG 120
 QY 121 TTTTAATCAGCCGGTCTTTATTTTACTTGTTTTGTCTAGTACGCTTGTCTCATATGTTATC 180
 Db 121 TTCATTCAGCCCATTTTATTCATGCTGCTCGTCTTTTGTGAGTTGCTGACGTATGTCATC 180
 QY 181 GTGAATATACGTTTATCCAAATTTATTTGATTTGTGACGAGCAGAAAAATTCGCTGCATG 240
 Db 181 GTGAATATTCGACTTTCCAAATTTATGATTTCTATACGAGCAGAAAAATTCGACGAATG 240
 QY 241 CTGATATCAGGAGTGTCTCTTAAAAATTCGGCTTGAATTTTCTATACCGGATGTGACATT 300
 Db 241 CTGATTAACGGGAATCTTTTGAATAATCGGTTTGAATTTATATATCCGGTGTAGCCGTTT 300
 QY 301 GAAATCGCAGAAATTTGAGGAAATCGGCATCATCGTCCAGGTTTATTTGCCAATTCATT 360
 Db 301 GAGATTCGCCGAATTCAGGGGAAATCGGAATCATCGTCCGGGCTGATTCGCCAATTCATT 360
 QY 361 CAGAAACAGAGTTTAAACAATTCGTTCCGAGACACGCTCTATTGACGCGAGCAGACTTT 420
 Db 361 CAAGACAGAGATTAACATTCATTTACGCTTGGAAAGTACGCTTTTATGACGCGAGCAACATT 420

Qy 421 GCTATCATGTTGTTTACTACTTAAT 446
Db 421 GTCAATTATGATGCTTACTACTTAAT 446

RESULT 8

AAH53465 standard; DNA; 453 BP.

AAH53465;

03-SEP-2001 (first entry)

S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2323.

Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination; endocarditis; ds.

Staphylococcus epidermidis.

MO200134809-A2.

17-MAY-2001.

09-NOV-2000; 2000WO-US030782.

09-NOV-1999; 99US-0164258P.

(GLAX) GLAXO GROUP LTD.

Kimerly WJ;

MPi; 2001-316495/33.

P-PSDB; AAG82615.

Nucleic acid encoding polypeptides from Staphylococcus epidermidis, useful for vaccinating against infections, e.g. endocarditis.

Claim 8; Page 617; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I) and (II) can have antibacterial activity and therefore be used in vaccination. The nucleic acids (I) may be used to produce the S. epidermidis polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention. N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464

Sequence 453 BP; 130 A; 60 C; 82 G; 181 T; 0 U; 0 Other;

Query Match 42.4%; Score 189.4; DB 4; Length 453;

Best Local Similarity 64.0%; Pred. No. 1.8e-47; Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 1 ATGTTGCGATGATTTATACATGCGACATATTTTGAAGTCTACTAGTTATTTT 60
Db 1 ATGATAGTTCGAAATTAATTTCTCTTATTCGAGGTCGTCGATGATTAATTT 60
Qy 61 GCGGAAAAAAGAGGATGTCGCGCAGGACTTGTGTACCGGAAATTTAGACTTGTG 120
Db 61 GCGGAAAAAAGAGGATGTCGCGCAGGACTTGTGTACCGGAAATTTAGACTTGTG 120

Db 61 GCTGAGAAATTTGGATTAATCCAGCAGGATGATCGTTCCAGGTTATTTAGCTTTGATT 120
Qy 121 TTATATGACCGCGCTTTATTTTACTGTTTGTGCTAGTAGACTTGCTCACTTAATGTTATC 180
Db 121 TTATATGACCGCGCTTTATTTTACTGTTTGTGCTAGTAGACTTGCTCACTTAATGTTATC 180
Qy 181 GTGAATATGCGTTTATCCAAATTTATGATTTTGTACGAGCAGCAAAAATTCGCTGCATG 240
Db 181 GTAGCAACGGTATTAAGTGGGTATTTTATATGATGAGAAATAATTCGCTGCATG 240
Qy 241 CTGATTAACAGGATCGCTCCATAAAAATCGCTTTGATTTTCTATPACCGGATTTGACATT 300
Db 241 ATACTGACGGGAATGATTAATTAATTTGATTTCTTTGATCCATGACCCCATTT 300
Qy 301 GAAATCCAGAAATTCGAGAAATCGCATCATCGTCCAGGTTTAAATTTGCCAATATCCATT 360
Db 301 GAAATGTTGAAATTTGAGATAGTGTGTGATCTCTGATATATTTGGAAATATACATT 360
Qy 361 CAGAAACAGGTTTAAACATTTACGTTGCGAAGCAGCTGCTATTGACGCGAGCACTTT 420
Db 361 CAAAACAGGTTGATGATTAACATTTACCTTTCAACAATGTTATTAACATGATTAACATAT 420
Qy 421 GCTATCATGTTGTTTACTACTTAAT 447
Db 421 ATCATCTTATTTTATATAGTTTAT 447

RESULT 9

ABN92203 standard; DNA; 459 BP.

ABN92203;

24-JUL-2002 (first entry)

Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1666.

Staphylococcus epidermidis; open reading frame; ORF; bacterial infection; antibacterial; gene therapy; gene; ds.

Staphylococcus epidermidis.

US6380370-B1.

30-APR-2002.

13-AUG-1998; 98US-00134001.

14-AUG-1997; 97US-0055779P.

08-NOV-1997; 97US-0064964P.

(GENO-) GENOME THERAPEUTICS CORP.

Doucette-Stamm LA, Bush D;

MPi; 2002-381255/41.

P-PSDB; ABP39658.

Novel isolated nucleic acid encoding a Staphylococcus epidermis polypeptide, useful for diagnosing and treating bacterial infections.

Disclosure; SEQ ID NO 1666; 267pp; English.

ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP35124 to ABP37960. The S. epidermidis sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly S. epidermidis infections. The sequences can be used to screen for compounds able to interfere with the S. epidermidis life cycle or inhibit S. epidermidis infection. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site

XX Sequence 459 BP; 135 A; 60 C; 82 G; 182 T; 0 U; 0 Other;
SQ

Query Match 42.4%; Score 189.4; DB 6; Length 459;
Best Local Similarity 64.0%; Pred. No. 1.8e-47;

Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```
OY 1 ATGTTGGATGAGTATTATACATCGCAGTATTTAGTGACTACTGATTTATTTT 60
DB 7 ATGATAGGTTGAGATTTATTTCTCTATTCGTAAGGTGTCGATCTGATTTGATTT 66
OY 61 GCGGAAAAACAGGATCGTCCCGCAGAGACTGTGTACCGGGATTTAGACTGTG 120
DB 67 GCTGGAATTTGGATTAATTCACAGGAGTACGTCGATTCAGTTATTTAGCTTTGAT 126
OY 121 TTTATACGCGGCTTTATTTTACTTGTGTTAGTACGCTCTCATTTATGTTATC 180
DB 127 TTTGATCAACCATCATGTTGTATCAAGTATTAATCATTTAGTTGCTTAATTTATC 186
OY 181 GTGAATACGCTTTATTCAGAAATTTATGATTTTGTACGACGAGAAATTCCTGCATG 240
DB 187 GTAGCAACGATTAAGTATGAGGTTATTTATTAATGAGAAATTCCTGCATG 246
OY 241 CTGATTAACAGGATCGTCTTAATAATCGGCTTTGATTTCTATACCCGATTTGACATTT 300
DB 247 ATACTGACGGGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 306
OY 301 GAAATCGCAGAAATTCAGAGAAATCGGATCATGTCGAGTTTATTCGCAATACCAT 360
DB 307 GAAATGTTGAAATTTAGGTTAGTGTGTCATTCCTGATTAATTTGCAATACCAT 366
OY 361 CAGAAACAGGTTTAAACATTAACGTTTCGAGACAGCTGCTATGAGCGAGACCTTT 420
DB 367 CAAAAACAGGTTGATGATCACTTCTACAAACATGTTATTAACATGATTAATCAAT 426
OY 421 GCTATCATGTTTGTCTTACTACTTAATT 447
DB 427 ATCATCTTATTTTATTAATGATTTTATT 453
```

RESULT 10

ADSO1974 standard; DNA; 459 BP.

ADSO1974;

04-NOV-2004 (first entry)

Staphylococcus epidermis polymnucleotide seqid 1269.

antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
recombinant expression vector; infection; computer readable medium;
computer based system; gene; ds.

Staphylococcus epidermidis.

US2004147734-A1.

29-JUL-2004.

01-DEC-2003; 2003US-00724972.

08-NOV-1997; 97US-0064964P.

13-AUG-1998; 98US-00134001.

29-NOV-1999; 99US-00450969.

(DOUC/) DOUCETTE-STAMM L.

(BUSH/) BUSH D.

Doucette-Stamm L, Bush D;
WPI, 2004-580138/56.
P-PSDB; ADS05746.

XX New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.

PS Claim 5; SEQ ID NO 1269; 741bp; English.

XX The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC a subject for S. epidermidis infection; a recombinant or substantially
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcal epidermidis bacterial
CC infection. This sequence encodes a S. epidermis protein of the invention.

XX Sequence 459 BP; 135 A; 60 C; 82 G; 182 T; 0 U; 0 Other;

Query Match 42.4%; Score 189.4; DB 13; Length 459;
Best Local Similarity 64.0%; Pred. No. 1.8e-47;

Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```
OY 1 ATGTTGGATGAGTATTATACATCGCAGTATTTAGTGACTACTGATTTATTTT 60
DB 7 ATGATAGGTTGAGATTTATTTCTCTATTCGTAAGGTGTCGATCTGATTTGATTT 66
OY 61 GCGGAAAAACAGGATCGTCCCGCAGAGACTGTGTACCGGGATTTAGACTGTG 120
DB 67 GCTGGAATTTGGATTAATTCACAGGAGTACGTCGATTCAGTTATTTAGCTTTGAT 126
OY 121 TTTATACGCGGCTTTATTTTACTTGTGTTAGTACGCTCTCATTTATGTTATC 180
DB 127 TTTGATCAACCATCATGTTGTATCAAGTATTAATCATTTAGTTGCTTAATTTATC 186
OY 181 GTGAATACGCTTTATTCAGAAATTTATGATTTTGTACGACGAGAAATTCCTGCATG 240
DB 187 GTAGCAACGATTAAGTATGAGGTTATTTATTAATGAGAAATTCCTGCATG 246
OY 241 CTGATTAACAGGATCGTCTTAATAATCGGCTTTGATTTCTATACCCGATTTGACATTT 300
DB 247 ATACTGACGGGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 306
OY 301 GAAATCGCAGAAATTCAGAGAAATCGGATCATGTCGAGTTTATTCGCAATACCAT 360
DB 307 GAAATGTTGAAATTTAGGTTAGTGTGTCATTCCTGATTAATTTGCAATACCAT 366
OY 361 CAGAAACAGGTTTAAACATTAACGTTTCGAGACAGCTGCTATGAGCGAGACCTTT 420
DB 367 CAAAAACAGGTTGATGATCACTTCTACAAACATGTTATTAACATGATTAATCAAT 426
OY 421 GCTATCATGTTTGTCTTACTACTTAATT 447
DB 427 ATCATCTTATTTTATTAATGATTTTATT 453
```

RESULT 11

AAH54379		standard; DNA; 3240 BP.
ID	AAH54379	standard; DNA; 3240 BP.
XX		
AC	AAH54379;	
XX		
DT	03-SEP-2001	(first entry)
XX		
DE	S. epidermidis genomic polynucleotide sequence SEQ ID NO:3743.	
XX		
KW	Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;	
XX	endocarditis; ds.	
OS	Staphylococcus epidermidis.	
XX		
FN	WO200134809-A2.	
PD	17-MAY-2001.	
XX		
PF	09-NOV-2000; 2000MO-US030782.	
XX		
PR	09-NOV-1999; 99US-0164258P.	
XX		
PA	(GLAXO) GLAXO GROUP LTD.	
XX		
PI	Kimmerly WJ;	
DR	WPI; 2001-316495/33.	
PT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,	
XX	useful for vaccinating against infections, e.g. endocarditis.	
PS	Claim 8; Page 1366-1367; 2188pp; English.	
CC	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides	
CC	(II), given in AAG81454 to AAG33120, from Staphylococcus epidermidis. (I)	
CC	and (II) can have antibacterial activity and therefore can be used in	
CC	vaccination. The nucleic acids (I) may be used to produce the S.	
CC	epidermidis polypeptides (II) via the production of vectors containing	
CC	them which are used to produce host cells which express the	
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be	
CC	used to vaccinate subjects and to raise antibodies against the bacteria.	
CC	The polypeptides may also be used to assay for other inhibitors of their	
CC	activity and therefore identify compounds that may be used for the	
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to	
CC	AAH5090 represent specifically claimed S. epidermidis genomic DNA	
CC	polynucleotide sequences from the present invention. AAH5091 to AAH5098	
CC	represent oligonucleotide sequences and primers which are used in the	
CC	exemplification of the present invention. N.B. The present invention	
CC	specifically claims all the polynucleotide sequences given in the	
CC	sequence listing of the present specification, however the sequence	
CC	listing only goes up to SEQ ID NO:4454 so even though sequences are given	
CC	in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present	
CC	for SEQ ID NO:4455 to 4464	
XX		
SEQ	Sequence 3240 BP; 1128 A; 464 C; 589 G; 1059 T; 0 U; 0 Other;	
	Query Match	40.7%; Score 182; DB 4; Length 3240;
	Best Local Similarity	65.5%; Pred. No. 6,6e-45;
	Matches 266; Conservative 0; Mismatches 140; Indels 0; Gaps 0	
OY	1 ATGTTGGATCGATTTATATACATCGCACTAATTNAGTGTAATACTACTAGTGAATTTT 60	
DB	2834 ATGATAGGTTCAGAATTATATATTTCTCTATTTCGTAGTGTCGTACTCAGTTGATATTT 289	
OY	61 GGCGAAAAAACAGGGATCTGCCGGCAGACCTGTGTACCGGAAATTTAGACTTGTG 120	
DB	2894 GCTGGAATTTGGGATTAATCATCAGAGGGTTAGTCTTCAGGTTAATTAGCTTTGATT 295	
OY	121 TTATATACGCGGCTTTATTTTAACTGTGTTTGCTAGTAGCTGCTCACTATGTTATC 180	
DB	2954 TTGTATCAACGCATCATCTGTTGTTATAGTAATTAATTAATTAATGTTCTTAATTTTATC 301	
OY	181 GTGAATACGGTTTATCCAAATTAATGATTTTGACGACGAGAAAATTCGCTGCATG 240	

Db	3014	GTAAGCAACGGTATTAGTAAGTGGGTTATTTTATATATGTTAAGAAAATTGGCTGCCATG	3072
Qy	241	CTGATTAACAGGAGATCGTCTTAAAAATGCGGCTTTGATTTCTATACCGGATTTACCATTT	300
Db	3074	ATACTGACGGGAATGGTATGATTAATTTATTTGATCTCTGTACCCATTACCCCATTT	3133
Qy	301	GAATTCGACAAATTTCCAGGAAATGGGCACTACTGTCGCCAGGTTTAATTCGCAATTCGCAAT	360
Db	3134	GAATTCGTTGAAGTTTCAGGTATAGGTGTGTCAATCTCTGTATATATTTGCGCAATACAAAT	3193
Qy	361	CAGAAACAAGGTTTATACCATTAACGTCGGGAAGCACGCTGCTATTGA	406
Db	3194	CAAAACAAGGTGATGTCATTACACTTTCTACAAACAATGTTATTAA	3239
RESULT 12			
ADB06111			
ID	ADB06111	standard; DNA; 480 BP.	
XX	AC	ADB06111;	
XX	AD		
XX	DT	20-NOV-2003 (first entry)	
XX	DE	Alloioococcus otitis antigenic protein encoding DNA SEQ ID NO:51.	
XX	KW	Alloioococcus otitidis; antigenic protein; immunogenic; immunisation;	
XX	KW	gene therapy; Gram-positive bacterium; infection; gene; ds.	
XX	OS	Alloioococcus otitidis.	
XX	PN	WO2003048304-A2.	
XX	PD	12-JUN-2003.	
XX	PE	25-NOV-2002; 2002WO-US036123.	
XX	PR	29-NOV-2001; 2001US-0333777P.	
XX	PR	18-NOV-2002; 2002US-0426742P.	
XX	PA	(AMHP) WYETH HOLDINGS CORP.	
XX	F1	Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;	
XX	DR	WPI; 2003-505284/47.	
XX	PT	P-PSDB; ADB06112.	
XX	PT	New Alloioococcus otitidis polynucleotides and polypeptides, useful for	
XX	PT	treating and diagnosing diseases, drug screening assays and monitoring of	
XX	PT	effects during drug clinical trials.	
XX	PS	Claim 7; SEQ ID NO 51; 1019bp; English.	
XX	XX	The present invention describes an isolated polynucleotide (I) of	
CC	CC	Alloioococcus otitidis genomic DNA, which encodes an antigenic protein.	
CC	CC	Alloioococcus otitidis is a Gram-positive bacterium. Also described: (1)	
CC	CC	an isolated polypeptide that is encoded by the polynucleotide (I); (2) an	
CC	CC	expression vector comprising the novel isolated polynucleotide (I), its	
CC	CC	host cell, transgene, transformed or infected with the vector of (2);	
CC	CC	(4) an antibody specific for the polypeptide of (1); (5) an immunogenic	
CC	CC	composition comprising the polypeptide, its complement, biological	
CC	CC	equivalent or fragment, or the polynucleotide that is comprised in the	
CC	CC	expression vector; (6) a pharmaceutical composition comprising the	
CC	CC	polypeptide of (1) and a carrier; (7) a protein chip comprising an array	
CC	CC	of the polypeptides of (1), their biological equivalent or fragment; (8)	
CC	CC	immunising against Alloioococcus otitidis by administering to a host the	
CC	CC	immunogenic composition; (9) detecting and/or identifying Alloioococcus	
CC	CC	otitidis in the biological sample; (10) a kit comprising a container	
CC	CC	containing the novel polynucleotide, its degenerate variant or fragment,	
CC	CC	or the antibody of (4); and (11) producing a polypeptide by culturing the	
CC	CC	genetically engineered host cell under conditions suitable to produce the	
CC	CC	polypeptide from the culture. (I) can be used in gene therapy. The	

CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting *Alloisococcus*
CC *oitidis*. The present sequence encodes an *Alloisococcus oitidis* antigen
CC protein from the present invention.

XX Sequence 480 BP; 113 A; 78 C; 99 G; 190 T; 0 U; 0 Other;

Query Match 14.6%; Score 65.2; DB 9; Length 480;
Best Local Similarity 49.5%; Pred. No. 1.8e-09;
Matches 200; Conservative 0; Mismatches 196; Indels 6; Gaps 1;

```
QY 9 ATCAGATTATATACATCGCACTAATTTAGTGTACTACTAGTATTTTGGCGAAA 68
DB 36 AACAGATTTTATATATGCTTGGCTTGAACATATCCATATAGTAGAAGCTT 95
QY 69 AACAGGATTCGCGCGGAGCACTGTGTACCGGGATTTAGAGCTGTGTTATCA 128
DB 96 TTTTGAAATTAATGCTGTGTATATGCCCCAGGATACCTAGCCCTTTATATGATCA 155
QY 129 GCCGCTTTATTTACTGTTTGTCTAGTGAAGCTTCACTTAATGTTATCGTGAATA 188
DB 156 GCCCTTGATGCTTCTTCAAGTATATTTATCTCAGCTAGCTAATTTGATTTATTA 215
QY 189 CGGTTATTCAAATTAATGATTTTGTACGAGCAGAAAATTCGCTGCATCTGATAC 248
DB 216 TGGGATTCGAATTCGTCATTTTATATGGAAGCGTTATTTGTTTAAATCTTGGT 275
QY 249 AGGATCGTCCCTAAATGCGCTTGA-----TTTCTATACCGATGTGCAATTTGA 302
DB 276 AGCTTTGGCATTAATAATGATTAATGAACTATTTCTTCTCTCTCTCTCTTGA 335
QY 303 AATCGAGATTTTCGAGAAATCGCATCATCGTCAGGTTTAAATGCCAATACCATTA 362
DB 336 TTCGTTATGTTTAGGGGATTTGGGCGAGTATGTTCCCGTTTACTGCGCAATATCTATG 395
QY 363 GAAACAAAGTTTAAACCATTAACGTTGCGAAGCAGCTGCTATTGCA 406
DB 396 ACGACAGTCACGCGATTAACCTTTGGCTCGTCATTTGGGATTTGA 439
```

RESULT 13
ADB06109
ID ADB06109 standard; DNA; 480 BP.

XX ADB06109;

XX 20-NOV-2003 (first entry)

XX *Alloisococcus oitidis* antigenic protein encoding DNA SEQ ID NO.49.

XX *Alloisococcus oitidis*; antigenic protein; immunogenic; immunisation;

XX gene therapy; Gram-positive bacterium; infection; gene; ds.

XX *Alloisococcus oitidis*.

XX MO2003048304-A2.

XX 12-JUN-2003.

XX 25-NOV-2002; 2002MO-US036123.

XX 29-NOV-2001; 2001US-033377P.

XX 18-NOV-2002; 2002US-0426742P.

XX (AMHP) WYETH HOLDINGS CORP.

XX Fletcher LD, Memichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

XX P-PDB; ADB06110.

PT New *Alloisococcus oitidis* polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.

XX Claim 7; SEQ ID NO 49; 1019pp; English.

XX The present invention describes an isolated polynucleotide (1) of
CC *Alloisococcus oitidis* genomic DNA, which encodes an antigenic protein.
CC *Alloisococcus oitidis* is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
CC expression vector comprising the novel isolated polynucleotide (1), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the array
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against *Alloisococcus oitidis* by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying *Alloisococcus*
CC *oitidis* in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting *Alloisococcus*
CC *oitidis*. The present sequence encodes an *Alloisococcus oitidis* antigen
CC protein from the present invention.

XX Sequence 480 BP; 113 A; 78 C; 99 G; 190 T; 0 U; 0 Other;

Query Match 14.6%; Score 65.2; DB 9; Length 480;
Best Local Similarity 49.5%; Pred. No. 1.8e-09;
Matches 200; Conservative 0; Mismatches 196; Indels 6; Gaps 1;

```
QY 9 ATCAGATTATATACATCGCACTAATTTAGTGTACTACTAGTATTTTGGCGAAA 68
DB 36 AACAGATTTTATATATGCTTGGCTTGAACATATCCATATAGTAGAAGCTT 95
QY 69 AACAGGATTCGCGCGGAGCACTGTGTACCGGGATTTAGAGCTGTGTTATCA 128
DB 96 TTTTGAAATTAATGCTGTGTATATGCCCCAGGATACCTAGCCCTTTATATGATCA 155
QY 129 GCCGCTTTATTTACTGTTTGTCTAGTGAAGCTTCACTTAATGTTATCGTGAATA 188
DB 156 GCCCTTGATGCTTCTTCAAGTATATTTATCTCAGCTAGCTAATTTGATTTATTA 215
QY 189 CGGTTATTCAAATTAATGATTTTGTACGAGCAGAAAATTCGCTGCATCTGATAC 248
DB 216 TGGGATTCGAATTCGTCATTTTATATGGAAGCGTTATTTGTTTAAATCTTGGT 275
QY 249 AGGATCGTCCCTAAATGCGCTTGA-----TTTCTATACCGATGTGCAATTTGA 302
DB 276 AGCTTTGGCATTAATAATGATTAATGAACTATTTCTTCTCTCTCTCTCTTGA 335
QY 303 AATCGAGATTTTCGAGAAATCGCATCATCGTCAGGTTTAAATGCCAATACCATTA 362
DB 336 TTCGTTATGTTTAGGGGATTTGGGCGAGTATGTTCCCGTTTACTGCGCAATATCTATG 395
QY 363 GAAACAAAGTTTAAACCATTAACGTTGCGAAGCAGCTGCTATTGCA 406
DB 396 ACGACAGTCACGCGATTAACCTTTGGCTCGTCATTTGGGATTTGA 439
```

RESULT 14
ADB12064_00/c
WP Sequence split into 18 fragments LOCUS ADB12064 Accession Adb12064
WP Fragment Name Begin End
WP ADB12064_00 1 110000

polypeptide, useful for diagnosing and treating bacterial infections.

Disclousure; SEQ ID NO 1489; 267bp; English.

ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP3124 to ABP37960. The *S. epidermidis* sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly *S. epidermidis* infections. The sequences can be used to screen for compounds able to interfere with the *S. epidermidis* life cycle or inhibit *S. epidermidis* infection. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site

Sequence 198 BP; 61 A; 51 C; 29 G; 57 T; 0 U; 0 Other;

Query Match 11.8%; Score 52.8; DB 6; Length 198;

Best Local Similarity 69.2%; Pred. No. 8e-06; Mismatches 32; Indels 0; Gaps 0;

Matches 72; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

1 ATGTCGATCAGATTATACATCGACTAATTTAGTGTACTACTCAGTTTAATTTT 60
104 ATGATAGCTTCAGAAATTAATTTCTCTATTCGTAGTGTGTCTACTCAGTTGATATT 45

61 GCGAATAAACAAGGATCGTCCGCGCAGGACTTGTGTACCGGG 104

44 GCTGAGAAATTTGGGATTAATCCAGCAGGGTTAGTGTTCAGG 1

Search completed: February 27, 2006, 07:05:18
Job time : 285.394 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28 ; Search time 489.476 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatcagattata.....tgttttctactactaatt 447

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications_NA_Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	447	100.0	447	US-10-484-605-2	Sequence 2, Appl1
2	447	100.0	447	US-10-789-164-2	Sequence 2, Appl1
3	447	100.0	636	US-10-484-605-6	Sequence 6, Appl1
4	289.2	64.7	450	US-09-974-300-2925	Sequence 2925, Ap
5	189.4	42.4	459	US-10-724-972A-1269	Sequence 1269, Ap
6	65.2	14.6	480	US-10-501-282-49	Sequence 49, Appl
7	65.2	14.6	480	US-10-501-282-51	Sequence 51, Appl
8	65.2	14.6	1754382	US-10-501-282-6651	Sequence 6651, Ap
9	39.6	8.9	3673778	US-10-312-841-1	Sequence 1, Appl1
10	39.6	8.4	5822	US-10-311-455-1069	Sequence 1069, Ap
11	36.2	8.1	26997	US-10-221-714A-474	Sequence 474, Appl
12	36	8.1	7195	US-10-239-676-30	Sequence 30, Appl
13	36	8.1	7195	US-10-240-453-40	Sequence 40, Appl
14	35.8	8.0	358	US-10-425-115-69390	Sequence 69390, A
15	35.2	7.9	2100	US-09-974-300-221	Sequence 221, Appl
16	35.2	7.9	4078	US-09-928-847B-43	Sequence 43, Appl
17	35	7.8	1347	US-10-282-122A-37840	Sequence 37840, A
18	35	7.8	11839	US-09-764-891-7628	Sequence 7628, Ap
19	35	7.8	3673778	US-10-312-841-2	Sequence 2, Appl1
20	34.4	7.7	534	US-10-139-667-6	Sequence 6, Appl1
21	34.4	7.7	637	US-10-027-632-225214	Sequence 225214, A
22	34.4	7.7	637	US-10-027-632-225215	Sequence 225215, A
23	34.4	7.7	637	US-10-027-632-225216	Sequence 225216, A

24	34.4	7.7	637	US-10-027-632-225215	Sequence 225215, A
25	34.4	7.7	5769	US-10-257-166-101	Sequence 101, Appl
26	34.4	7.7	51259	US-10-374-077-209	Sequence 209, Appl
27	34.2	7.7	1344	US-10-472-928-2631	Sequence 2631, Appl
28	34.2	7.7	5273	US-10-311-455-847	Sequence 847, Appl
29	34.2	7.7	6052	US-10-311-455-390	Sequence 390, Appl
30	34.2	7.7	12666	US-08-961-527-137	Sequence 137, Appl
31	34.2	7.7	12666	US-10-158-844-137	Sequence 137, Appl
32	34.2	7.7	277616	US-10-367-094-83	Sequence 83, Appl1
33	34.2	7.7	216258	US-10-472-928-4979	Sequence 4979, Ap
34	34	7.6	1905	US-10-741-849-6083	Sequence 6083, Ap
35	34	7.6	14023	US-10-311-455-2078	Sequence 2078, Appl
36	34	7.6	34769	US-10-221-714A-502	Sequence 502, Appl
37	33.8	7.6	7046	US-10-311-455-2090	Sequence 2090, Ap
38	33.8	7.6	110079	US-10-175-523-96	Sequence 96, Appl1
39	33.8	7.6	110079	US-11-099-266-96	Sequence 96, Appl1
40	33.6	7.5	389	US-09-925-065A-307190	Sequence 307190, A
41	33.6	7.5	7173	US-11-097-143-22021	Sequence 22021, A
42	33.6	7.5	8303	US-10-311-455-395	Sequence 395, Appl
43	33.6	7.5	96109	US-10-672-787-35	Sequence 35, Appl
44	33.4	7.5	423	US-10-424-599-25187	Sequence 25187, A
45	33.4	7.5	460	US-09-918-995-26253	Sequence 26253, A

ALIGNMENTS

RESULT 1
US-10-484-605-2
Sequence 2, Application US/10484605
Publication No. US20040253704A1
GENERAL INFORMATION:
APPLICANT: Sung, Moon-Hee
APPLICANT: Hong, Seung-Pyo
APPLICANT: Lee, Jong-Su
APPLICANT: Jung, Chang-Min
APPLICANT: Kim, Chul-Joong
APPLICANT: Soda, Kenji
APPLICANT: Ashuchi, Makoto
TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING P98BCA, THE GENE CODING
TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
TITLE OF INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
FILE REFERENCE: 4240-101
CURRENT APPLICATION NUMBER: US/10/484,605
PRIOR FILING DATE: 2004-01-20
PRIOR APPLICATION NUMBER: PCT/KR02/01522
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: KR 2001-48373
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2
LENGTH: 447
TYPE: DNA
ORGANISM: Bacillus subtilis
US-10-484-605-2

Query Match 100.0%; Score 447; DB 8; Length 447;

Best Local Similarity 100.0%; Pred. No. 1.6e-120;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGTTGGATCAGATTATTAACATCGACATTAATTTAGGTGCTACTGCTCACTTAATTTT	60
DB	1	ATGTTGGATCAGATTATTAACATCGACATTAATTTAGGTGCTACTGCTCACTTAATTTT	60
QY	61	GGGGAAGAAACAGGGATCGGCGGAGGACTGTTGTACCGGATATTAGGACTGTG	120
DB	61	GGGGAAGAAACAGGGATCGGCGGAGGACTGTTGTACCGGATATTAGGACTGTG	120
QY	121	TTTAATCAGCGGCTTTTATTTACTGTTTGTAGTAGAGCTTGCTCACTTAATTTT	180
DB	121	TTTAATCAGCGGCTTTTATTTACTGTTTGTAGTAGAGCTTGCTCACTTAATTTT	180

QY	181	GTGAATAACGGTTATCCAAATTTATGATTTTGTACGAGACGAGAAAATTTGGTCGCATG	240
Db	181	GTGAATAACGGTTATCCAAATTTATGATTTTGTACGAGACGAGAAAATTTGGTCGCATG	240
QY	241	CTGATTAACAGGGATCGTCTTAAAAAATCGCGTTTGAATTTTCTAATCCGAGATTGTAACATTT	300
Db	241	CTGATTAACAGGGATCGTCTTAAAAAATCGCGTTTGAATTTTCTAATCCGAGATTGTAACATTT	300
QY	301	GAATTCGAGAAATTTGAGGAATCGGCATCATCGTGTACGAGGTTTAATTTGCCAATACCATTT	360
Db	301	GAATTCGAGAAATTTGAGGAATCGGCATCATCGTGTACGAGGTTTAATTTGCCAATACCATTT	360
QY	361	CAGAAACAGGTTTAACCAATTACGTTGCGAAGCAAGCTGTATTGTAGCGGAGCAACCTTT	420
Db	361	CAGAAACAGGTTTAACCAATTACGTTGCGAAGCAAGCTGTATTGTAGCGGAGCAACCTTT	420
QY	421	GCTATCATGTTGTTTACTACTTAATT	447
Db	421	GCTATCATGTTGTTTACTACTTAATT	447

```

RESULT 2
US-10-789-164-2
: Sequence 2, Application US/10789164
: Publication No. US20050191720A1
: GENERAL INFORMATION:
: APPLICANT: SUNG, M.H. et al.
: TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENOME
: FILE REFERENCE: P1574
: CURRENT APPLICATION NUMBER: US/10/789,164
: CURRENT FILING DATE: 2004-02-27
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: KopatecIntn 1.71
: SEQ ID NO 2
: LENGTH: 447
: TYPE: DNA
: ORGANISM: Bacillus subtilis
: US-10-789-164-2

```

Query Match	100.0%	Score 447	DB 9	Length 447
Best Local Similarity	100.0%	Pred. No. 1,6e-120		
Matches 447	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	ATGTCGGATCAGATTATATACATCGCACTAATTTTAGGTACTACTAGTTTAAATTTT	60	
Db	1	ATGTCGGATCAGATTATATACATCGCACTAATTTTAGGTACTACTAGTTTAAATTTT	60	
QY	61	GCGGAAAAAAGAGGATGTCGCGGACGACTGTGTGACCGGGATATTAGGACTGTG	120	
Db	61	GCGGAAAAAAGAGGATGTCGCGGACGACTGTGTGACCGGGATATTAGGACTGTG	120	
QY	121	TTTATACGCGGCTTTATATTTTACTGTGTTGCTAGTAGAGCTTGCTCACTTAATGTAAC	180	
Db	121	TTTATACGCGGCTTTATATTTTACTGTGTTGCTAGTAGAGCTTGCTCACTTAATGTAAC	180	
QY	181	GTGAATAATCGGTTTATCCAAATTTATGATTTTGTACGGAACCCAGAAAAATTGGCTGCATG	240	
Db	181	GTGAATAATCGGTTTATCCAAATTTATGATTTTGTACGGAACCCAGAAAAATTGGCTGCATG	240	
QY	241	CTGATTAACAGGATCGTCTTAAAAATCCGCTTTGATTTTCTATATCCGATGTACAAATTT	300	
Db	241	CTGATTAACAGGATCGTCTTAAAAATCCGCTTTGATTTTCTATATCCGATGTACAAATTT	300	
QY	301	GAATTCGCAGAATTTCCAGGAATGCGCATCATCTGTGCCAGGTTTATTTGCCAATACCAAT	360	
Db	301	GAATTCGCAGAATTTCCAGGAATGCGCATCATCTGTGCCAGGTTTATTTGCCAATACCAAT	360	
QY	361	CAGAAACAAAGGTTTAAACCATACGTTGGAAGACAGCGCTGATTAGAGGAGACGACCTTT	420	
Db	361	CAGAAACAAAGGTTTAAACCATACGTTGGAAGACAGCGCTGATTAGAGGAGACGACCTTT	420	
QY	421	GCTATCATGTTGTTTACTACTTAATTT	447	

Db 421 GCTATCATGTTGTTACTACTTAAT 447

```

RESULT 3
US-10-484-605-6
Sequence 6, Application US/10484605
Publication No. US20040253704A1
GENERAL INFORMATION:
APPLICANT: Sung, Moon-Hee
APPLICANT: Hong, Seung-Pyo
APPLICANT: Lee, Jong-Su
APPLICANT: Jung, Chang-Min
APPLICANT: Kim, Chul-Joong
APPLICANT: Soda, Kenji
APPLICANT: Ashiuchi, Makoto
TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING PSBACA, THE GENE CODING
TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
TITLE OF INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
FILE REFERENCE: 4240-101
CURRENT APPLICATION NUMBER: US/10/484,605
CURRENT FILING DATE: 2004-01-20
PRIOR APPLICATION NUMBER: PCT/KR02/01522
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: KR 2001-48373
PRIOR FILING DATE: 2001-08-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 6536
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Construct
US-10-484-605-6

```

Query Match	100.0%	Score 447	DB 8	Length 6536
Best Local Similarity	100.0%	Pred. No. 5	Be-120	
Matches 447	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	ANGTTCGGATCAGATTATATACATCGCACTAATTTTAGGTGTACTATCTCAGTTTAATTTTT	60	
Db	1423	ATGTTCCGATCAGATTATATACATCGCACTAATTTTAGGTGTACTACTCAGTTTAATTTTT	1482	
QY	61	GGGAAAAAACAAGGATGTGCGCGGACGACCTGTGTATACCGGGATATTTAGACTTG	120	
Db	1483	GGGAAAAAACAAGGATGTGCGCGGACGACTTGTGTATACCGGATATTTAGACTTG	1542	
QY	121	TTTATACAGCCGGTCTTATTTTACTTTTGCTAGTAGCTTGCTCATTATCTATC	180	
Db	1543	TTTATACAGCCGGTCTTATTTTACTTTTGCTAGTAGCTTGCTCATTATCTATC	1602	
QY	181	GTGAATATACGGTTTATCCAAATTTATGATTTTGTACGGACGAGAAAATTTGGCTCCGAT	240	
Db	1603	GTGAATATACGGTTTATCCAAATTTATGATTTTGTACGGACGAGAAAATTTGGCTCCGAT	1662	
QY	241	CTGATTAACAGGGATCGTCTTAAAAATCGCGTTTGAATTTTCTAATCCGATGTGACATTT	300	
Db	1663	CTGATTAACAGGGATCGTCTTAAAAATCGCGTTTGAATTTTCTAATCCGATGTGACATTT	1722	
QY	301	GAATTCGAGAAATTTGAGGAATGGGCAATCATCGCCAGGTTTAATTTGCCAATTCACAT	360	
Db	1723	GAATTCGAGAAATTTGAGGAATGGGCAATCATCGCCAGGTTTAATTTGCCAATTCACAT	1782	
QY	361	CAGAAACAGGTTTAAACCATTTACGTTGGAAGACAGCTGCTATTGAGCGGAGCGACCTTT	420	
Db	1783	CAGAAACAGGTTTAAACCATTTACGTTGGAAGACAGCTGCTATTGAGCGGAGCGACCTTT	1842	
QY	421	GCTATCATGTTGTTTACTACTTAATT 447		
Db	1843	GCTATCATGTTGTTTACTACTTAATT 1869		

RESULT 4
US-09-974-300-2925
Sequence 2925, Application US/09974300
Patent No. US2002016721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2925.
LENGTH: 450
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-2925

Query Match 64.7%; Score 289.2; DB 3; Length 450;
Best Local Similarity 78.0%; Pred. No. 3.2e-74;
Matches 348; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 ATGTTGGATCGAGTTTATACATGCACTAATTTAGGTGTAAGTCTGAGTTAATTTT 60
DB 1 ATGTTGGATCGAGTTTATACATGCACTAATTTAGGTGTAAGTCTGAGTTAATTTT 60
QY 61 GCGAAAAACAGGAGATCGTCCGCGAGGAGCTGTGTGACCGGATATTAGACTGTG 120
DB 61 GCGAAAAACAGGAGATCGTCCGCGAGGAGCTGTGTGACCGGATATTAGACTGTG 120
QY 121 TTTAATCAGCGGTCTTATTTTCTGTTTGTGCTAGTGAAGCTGTCTACTATGTAAC 180
DB 121 TTTAATCAGCGGTCTTATTTTCTGTTTGTGCTAGTGAAGCTGTCTACTATGTAAC 180
QY 121 TTTAATCAGCGGTCTTATTTTCTGTTTGTGCTAGTGAAGCTGTCTACTATGTAAC 180
DB 121 TTTAATCAGCGGTCTTATTTTCTGTTTGTGCTAGTGAAGCTGTCTACTATGTAAC 180
QY 181 GTGAATACGGATTTATCCAAATTTATGTTTGTACGAGCAGAAAAATTCGTCGCATG 240
DB 181 GTGAATACGGATTTATCCAAATTTATGTTTGTACGAGCAGAAAAATTCGTCGCATG 240
QY 241 CTGATACAGGAGATCGTCTAAATAATCGGTTGATTTCTATACCCGATGACCATTT 300
DB 241 CTGATACAGGAGATCGTCTAAATAATCGGTTGATTTCTATACCCGATGACCATTT 300
QY 301 GAAATCGAGAAATTTGAGGAAATCGGCATCATGTCAGAGTTAAATGCGAATACCAT 360
DB 301 GAAATCGAGAAATTTGAGGAAATCGGCATCATGTCAGAGTTAAATGCGAATACCAT 360
QY 361 CAGAAACAGGTTTAAACATTAAGTTCGAGACGACCTGATATGAGGAGGACCTTT 420
DB 361 CAGAAACAGGTTTAAACATTAAGTTCGAGACGACCTGATATGAGGAGGACCTTT 420
QY 421 GCTATCATGTTTGTACTACTAAT 446
DB 421 GCTATCATGTTTGTACTACTAAT 446

RESULT 5
US-10-724-972A-1269
Sequence 1269, Application US/10724972A
Publication No. US2004014773A1
GENERAL INFORMATION:
APPLICANT: Doucette-Stamm, Lynn
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PTH03-16
CURRENT APPLICATION NUMBER: US/10/724,972A
CURRENT FILING DATE: 2003-12-01

PRIOR APPLICATION NUMBER: 09/450,969
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 7544
SEQ ID NO 1269
LENGTH: 459
TYPE: DNA
ORGANISM: S.epidermidis
US-10-724-972A-1269

Query Match 42.4%; Score 189.4; DB 7; Length 459;
Best Local Similarity 64.0%; Pred. No. 6.2e-45;
Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 1 ATGTTGGATCGAGTTTATACATGCACTAATTTAGGTGTAAGTCTGAGTTAATTTT 60
DB 7 ATGTTGGATCGAGTTTATACATGCACTAATTTAGGTGTAAGTCTGAGTTAATTTT 66
QY 61 GCGAAAAACAGGAGATCGTCCGCGAGGAGCTGTGTGACCGGATATTAGACTGTG 120
DB 67 GCTGAGAAATTTGGATTAATTCAGCAGGAGTGAAGCTGTCCAGGTTATTTAGCTTGAAT 126
QY 121 TTTAATCAGCGGTCTTATTTTCTGTTTGTGCTAGTGAAGCTGTCTACTATGTAAC 180
DB 127 TTTAATCAGCGGTCTTATTTTCTGTTTGTGCTAGTGAAGCTGTCTACTATGTAAC 186
QY 181 GTGAATACGGATTTATCCAAATTTATGTTTGTACGAGCAGAAAAATTCGTCGCATG 240
DB 187 GTGAATACGGATTTATCCAAATTTATGTTTGTACGAGCAGAAAAATTCGTCGCATG 246
QY 241 CTGATACAGGAGATCGTCTAAATAATCGGTTGATTTCTATACCCGATGACCATTT 300
DB 247 ATATGACGAGGAAATGATTAATTTATGTTTGTATGATTCCTGATACCATGACCCCATTT 306
QY 301 GAAATCGAGAAATTTGAGGAAATCGGCATCATGTCAGAGTTAAATGCGAATACCAT 360
DB 307 GAAATCGAGAAATTTGAGGAAATCGGCATCATGTCAGAGTTAAATGCGAATACCAT 366
QY 361 CAGAAACAGGTTTAAACATTAAGTTCGAGACGACCTGATATGAGGAGGACGACCTTT 420
DB 367 CAGAAACAGGTTTAAACATTAAGTTCGAGACGACCTGATATGAGGAGGACGACCTTT 426
QY 421 GCTATCATGTTTGTACTACTAAT 447
DB 427 ATCATCTATTTTATATAGTTTATTT 453

RESULT 6
US-10-501-282-49
Sequence 49, Application US/10501282
Publication No. US20050203280A1
GENERAL INFORMATION:
APPLICANT: MICHAEL, JOHN CALHOUN
APPLICANT: ZAGURSKY, ROBERT JOHN
APPLICANT: RUSSELL, DAVID PARRISH
APPLICANT: FLETCHER, LEAH DIANE
TITLE OF INVENTION: ALLOTOCOCUS OTITIDIS OPEN READING FRAMES (ORFs) ENCODING
TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
FILE REFERENCE: AM100780 L2
CURRENT APPLICATION NUMBER: US/10/501,282
PRIOR FILING DATE: 2004-07-09
PRIOR APPLICATION NUMBER: 60/333,777
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 60/426,742
PRIOR FILING DATE: 2002-11-18
PRIOR APPLICATION NUMBER: PCT/US02/36123
PRIOR FILING DATE: 2002-11-25
NUMBER OF SEQ ID NOS: 6653

129 GCCGCTTTATTTACTTTGTTTCTAGTACGCTTCTCACTTATGTTATCGTAATA 188
22110 GCCCTTGATGCTTCTCAGTATATTTATCTCACTGCTAGTATTTGATTTGTAATA 22051
189 CGGTTATCCAAATTTATGATTTTGTACGAGCAGAAATTCGTCGATGCTGATAC 248
22050 TGGGATTCGAAATTCGATTTTATGAGGAGGCTTATTTGTTTAAATCTTGGT 21991
249 AGGATCTCTTAAATTCGCTTTGA-----TTTCTATACCGGATTTGACATTTGA 302
21990 AGCTTTGCAATTAATATGATTTATGAACTATTTCTACTCTCTCCCTTAGA 21931
303 AATGCAGAATTTGAGAGATCGCATCATCTGCGAGTTTAAATTCGCAATACCATTA 362
21930 TTCCGTTATGTTTACGGGGGATTTGGGAGTACGCTCCGCTTACCGCAATATCTATCG 21871
353 GAAAGAGGTTTAACTATGCTTGGGAAGCAGCGCTGCTATTGA 406
21870 AGCAGATCATGCGGATTAATTTGCTGCTGCTATTTGGAGTTGA 21827

RESULT 9
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/NO
; CURRENT APPLICATION NUMBER: US/10/312, 841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
; US-10-312-841-1

Query Match 8.9%; Score 39.6; DB 6; Length 3673778;
Best Local Similarity 44.0%; Pred. No. 38;
Matches 168; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

1 ATGTTGATCAGATTATTAATCATGCACTAATTTTATGCTTACTCACTTAAATTTT 60
1316265 ATTAATGTTTGAATTTAATTAAGTTTATTTAAGTTTAAATGATTAATAGTTT 1316324
61 GCGGAAAAAAGAGGATCGCGCCGAGAGCTGTTGTACCGGGAATTTAGACTTTG 120
1316325 ATTAATGTTTGAATTTAATTAAGTTTATTTAATGTTTAAATGATTAATAGTTT 1316384
121 TTTAATCAGCGCTTATTTACTTGTCTAGTAGTGAAGTGTCTCACTTATGTTATC 180
1316385 ATTTGAGTTTGAATTTATTTAGTTTATAGTAGTTTGAAGTATTTATTTAT 1316444
181 GTGAAATACGTTTATTCAAATTTATGATTTTGTACGAGCAGCAAAATTCGCTGCCAG 240
1316445 TTTAATGAGTTTGTGATTTATTAATTAATGTTTATTTAATGAGTTTATTAATAGT 1316504
241 CTGATACAGGAGATCGCTTAAATTCGCTTATTTCTATACCGGATTTGACATTT 300
1316505 TTTAATGAAAGTTTGTGATTTATTAATTAATTAATGAGTTTGAATTTATTAAGT 1316564
301 GAATCGCAGAAATTTGAGAGATCGCATCATCTGCGAGTTTAAATTCGCAATACCAT 360
1316565 TTTATTTGAGTTTGTGAGATTTATTTATTTATTTAATGAAGTTTGTGATTTATTAAGT 1316624
361 CAGAAACAGGTTTAAACATTA 382

1316625 TTTATGAGGTTTGTGAGATTA 1316646

RESULT 10
US-10-311-455-1069
; Sequence 1069, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENEROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311, 455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1069
; LENGTH: 5822
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-311-455-1069

Query Match 8.4%; Score 37.4; DB 6; Length 5822;
Best Local Similarity 57.1%; Pred. No. 8.1;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

106 TATTAGGACTGTGTTTATCATGCGGCTTATTTACTTGTGTTGCTAGTACGCTTG 165
4775 TATGTTGAAGTTGTTTAAATATAGTTTGTGTTTATTTGTTAGGTATTTTGT 4834
166 CTCACCTATGTTATCGTGAAATACGTTTATTCAAATTTATGATTTTGTACGAGCGAG 224
4835 TTTATTTAGTTTATGAGAAATTTTAAATTTATTTATTTAGTTTATTTTAAAGTAG 4893

RESULT 11
US-10-221-714A-474
; Sequence 474, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENEROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Cytosine Methylation
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221, 714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 474
; LENGTH: 26997

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-221-714A-474
Query Match      8.1%; Score 36.2; DB 7; Length 26997;
Best Local Similarity 52.3%; Pred. No. 37;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 67 AAACAGGAGATCGTCCGACGAGACTTGTGACCGGATTTAGACCTGTGTTAAT 126
DB 15921 AAGAAAGGAGAGCGGTGTGATTCGAAAGTTGAAATGATGTTTGTGTTATA 15980
OY 127 CAGCCGCTTATTTACTGTTGCTAGTACGCTTGCCTACTATGTTATCGTGAAA 186
DB 15981 TAACGTGAATTAGATTATGTTGTTGAAGATGTTGTTGTTGTTGTTGTTAAG 16040
OY 187 TACGCTTATCCAAATTTATGATTTTGTACGGA 219
DB 16041 TTAATTATAAGATATTTAAGTTGTTGCTGA 16073

RESULT 12
US-10-239-676-30
; Sequence 30, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT FILING DATE: 2002-09-24
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 30
; LENGTH: 7195
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-239-676-30
Query Match      8.1%; Score 36; DB 5; Length 7195;
Best Local Similarity 54.5%; Pred. No. 23;
Matches 72; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY 84 GGCAGGACTGTGTAACGGGATTTAGAGACTGTGTTAATCAGCCGCTTTATTT 143
DB 4814 GGCGTAGATTATCGATTGTTTATGCTGTTTTTTTATTTAGATTATGTTATTTT 4873
OY 144 ACTTGTTCAGAGAGCTTCTCACTATGTTATCGGAATACGTTTATCCAAAT 203
DB 4874 ATTAGTTATGTTGATTATTTAATTATTTAAGTTAGTTAATTTGAAATTTTGA 4933
OY 204 TATGATTTGTA 215
DB 4934 TGTAAGTTAGTA 4945

RESULT 13
US-10-240-453-40
```

```

; Sequence 40, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: With Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: with DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR FILING DATE: 2000-04-07
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 40
; LENGTH: 7195
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-240-453-40
Query Match      8.1%; Score 36; DB 6; Length 7195;
Best Local Similarity 54.5%; Pred. No. 23;
Matches 72; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY 84 GGCAGGACTGTGTAACGGGATTTAGAGACTGTGTTAATCAGCCGCTTTATTT 143
DB 4814 GGCGTAGATTATCGATTGTTTATGCTGTTTTTTTATTTAGATTATGTTATTTT 4873
OY 144 ACTTGTTCAGAGAGCTTCTCACTATGTTATGTTAATCGGAATACGTTTATCCAAAT 203
DB 4874 ATTAGTTATGTTGATTATTTAATTATTTAAGTTAGTTAATTTGAAATTTTGA 4933
OY 204 TATGATTTGTA 215
DB 4934 TGTAAGTTAGTA 4945

RESULT 14
US-10-425-115-69390/c
; Sequence 69390, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21 (53222)B
; CURRENT APPLICATION NUMBER: US/10/425.115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 69390
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MNT4577_163278C.1
US-10-425-115-69390
Query Match      8.0%; Score 35.8; DB 8; Length 358;
```


Best Local Similarity 56.3%; Pred. No. 6.4;
Matches 67; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 48 CAGTTAAATTTTGGGAAAAAAGAGATGTCGCGCAGACTTTGTAACGGGATA 107
166 CAGCTTTTTCAGTGGGATTAAGTAAACGGGGGGGGGGCTTTCCCTGGGGAGG 107
Db 108 TTTAGACTTGTGTTAATCAGCCCGCTTTTATTACTGTTTGTCTAGTGAAGCTGC 166
106 GTTATTAACACCTGTAAATGGGGCGTTTATTTTATTTTGAAGATTAACTGGC 48

RESULT 15

US-09-974-300-221
; Sequence 221, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 221
; LENGTH: 2100
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-221

Query Match 7.9%; Score 35.2; DB 3; Length 2100;
Best Local Similarity 52.8%; Pred. No. 22;
Matches 76; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 203 TTATGATTTTGTACGAGCAGAAAATTCGCTGCATGATTAACAGGATGCTCTAA 262
1492 TTACCAATTTTAACTGTCTTTGCCCCATATGACTGCGAGAAAAGAGATGCTTCC 1551
QY 263 AATCGCGTTTGTATTTCTATACCGATTGTATCAATTTGAATCGAGAATTTGAGGAA 322
1552 AATGCCCTGCGCCCTCCGAAAGAGTTGAAGCGCTTGAAGAAAAGCAGAAATTCATCA 1611
Db 323 TCGGCATCATCGTCCAGGTTTAA 346
1612 TTCAGGTGATGAACCTGCCTTGA 1635

Search completed: February 27, 2006, 07:58:32.
Job time : 498.476 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 269.374 Seconds
(without alignments)
3538.457 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatcagattata.....tgccttcttactactaatc 447

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New:
1: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
2: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq1:*
8: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq2:*
10: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq2:*
11: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq3:*
12: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq4:*
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	44.7	100.0	447	US-10-530-083-2	Sequence 2, Appl1
2	189.4	42.4	453	US-10-793-626-2323	Sequence 2323, Ap
3	182	40.7	3240	US-10-793-626-3743	Sequence 3743, Ap
4	45.6	10.2	3029	US-10-793-626-4197	Sequence 4197, Ap
5	37.8	8.5	168753	US-11-181-234-1	Sequence 1, Appl1
6	33.6	7.5	389	US-09-925-065A-307190	Sequence 307190,
7	33.4	7.5	498	US-09-925-065A-297021	Sequence 297021,
8	33.4	7.5	528	US-09-925-065A-780420	Sequence 780420,
9	33.4	7.5	530	US-09-925-065A-804269	Sequence 804269,
10	33.4	7.5	569	US-09-925-065A-547674	Sequence 547674,
11	33.4	7.5	569	US-09-925-065A-547675	Sequence 547675,
12	33.4	7.5	611	US-09-925-065A-418523	Sequence 418523,
13	33.4	7.5	611	US-09-925-065A-418524	Sequence 418524,
14	33.4	7.5	1524	US-09-925-065A-692556	Sequence 692556,
15	33.4	7.5	1524	US-09-925-065A-692557	Sequence 692557,
16	33.4	7.5	1524	US-09-925-065A-692558	Sequence 692558,
17	33.4	7.5	1524	US-09-925-065A-692559	Sequence 692559,
18	33.4	7.5	1524	US-09-925-065A-692560	Sequence 692560,
19	33.4	7.5	1524	US-09-925-065A-692561	Sequence 692561,
20	33.4	7.5	2336	US-09-925-065A-674512	Sequence 674512,

ALIGNMENTS

RESULT 1	US-10-530-083-2	US-09-925-065A-674513	Sequence 674513,
1	Sequence 2, Application US/10530083	US-09-925-065A-709850	Sequence 709850,
2	Publication No. US20050249752A1	US-09-925-065A-709851	Sequence 709851,
3	GENERAL INFORMATION:	US-09-925-065A-709852	Sequence 709852,
4	APPLICANT: Sung, Moon-hee	US-09-925-065A-709853	Sequence 709853,
5	APPLICANT: Poo, Ha Ryoung	US-09-925-065A-709854	Sequence 709854,
6	APPLICANT: Lee, Jong-Seo	US-10-240-708-20	Sequence 20, Appl
7	APPLICANT: Jung, Chang-Min	US-09-925-065A-325389	Sequence 325389,
8	APPLICANT: Hong, Seong-Pyo	US-10-750-185-38179	Sequence 38179, A
9	APPLICANT: Kim, Chul-Joong	US-10-750-623-18179	Sequence 38179, A
10	APPLICANT: Park, Sue-nie	US-09-925-065A-588494	Sequence 588494,
11	APPLICANT: Pyo, Hyun-mi	US-09-925-065A-835481	Sequence 835481,
12	TITLE OF INVENTION: VECTOR FOR ANTI-HPV VACCINE AND TRANSFORMED MICROORGANISM BY THE	US-09-925-065A-897815	Sequence 897815,
13	FILE REFERENCE: 4240-119	US-09-925-065A-500127	Sequence 500127,
14	CURRENT APPLICATION NUMBER: US/10/530,083	US-09-925-065A-500128	Sequence 500128,
15	CURRENT FILING DATE: 2005-04-01	US-09-925-065A-96929	Sequence 96929, A
16	PRIOR FILING DATE: 2002-10-17	US-09-925-065A-483171	Sequence 483171,
17	NUMBER OF SEQ ID NOS: 11	US-09-925-065A-588495	Sequence 588495,
18	SOFTWARE: PatentIn version 3.2	US-09-925-065A-292480	Sequence 292480,
19	SEQ ID NO 2	US-09-925-065A-939567	Sequence 939567,
20	LENGTH: 447	US-09-925-065A-939568	Sequence 939568,
21	ORGANISM: Bacillus subtilis	US-11-117-187-212	Sequence 212, App
22	TYPE: DNA		
23	Query Match	100.0%; Score 447; DB 8; Length 447;	
24	Best local Similarity	100.0%; Pred. No. 7.1e-115;	
25	Matches	447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
26	US-10-530-083-2		
27	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
28	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
29	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
30	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
31	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
32	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
33	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
34	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
35	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
36	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
37	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
38	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
39	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
40	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
41	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
42	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
43	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
44	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		
45	1 ATGTTGAGATCAGATTATATACATGCACTAATTTTGGTGTACTACTGATTAATTTT		

OY 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
DB 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
OY 241 CTGATTAACAGGAGATGCTCTTAAATTCGCGTTTGAATTTTCTATACCGGATTTGATTAATTT 300
DB 241 CTGATTAACAGGAGATGCTCTTAAATTCGCGTTTGAATTTTCTATACCGGATTTGATTAATTT 300
OY 301 GAAATCGAGAAATTTTCGAGGATTCGGCATTCGCGATTCGCAAGTTTAATTCGCAATTCATTT 360
DB 301 GAAATCGAGAAATTTTCGAGGATTCGGCATTCGCGATTCGCGATTTAATTCGCAATTCATTT 360
OY 361 CAGAAACAGGTTTAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 420
DB 361 CAGAAACAGGTTTAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 420
OY 421 GCTATCATGTTTGTCTTACTTAATTT 447
DB 421 GCTATCATGTTTGTCTTACTTAATTT 447

RESULT 2
US-10-793-626-2323

; Sequence 2323, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2323
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2323

Query Match 42.4%; Score 189.4; DB 8; Length 453;
Best Local Similarity 64.0%; Pred. No. 4.5e-43;
Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

OY 1 ATGTTGGATCGATTTATACATCGACATTAATTTAGGTGATCTACGATTAATTTT 60
DB 1 ATGATAGGTTCCGAATTTATTTCTCTTATTCGAGGTTGTCGATCTCAAGTTGATTTT 60
OY 61 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGTACCGGATTTTATGAGCTTGTG 120
DB 61 GCTGAGAAATTTGGATTAATCCAGAGGTTAGTGTTCCAGGTTATTTAGCTTTGATTT 120
OY 121 TTTAATCAGCGGCTTTATTTTACTTGTGTTGCTAGTAGAGCTTCTCATTTAGTTATC 180
DB 121 TTTGATCAACCGATCATGTTGTTATTCAGATTAATCATTTAGTCTTAACTTAATTTATC 180
OY 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
DB 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
OY 241 CTGATTAACAGGAGATGCTCTTAAATTCGCGTTTGAATTTTCTATACCGGATTTGATTAATTT 300
DB 241 ATACTGACGGGATGATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 300
OY 301 GAAATCGAGAAATTTTCGAGGATTCGGCATTCGCGATTCGCAAGTTTAATTCGCAATTCATTT 360
DB 301 GAAATCGAGAAATTTTCGAGGATTCGGCATTCGCGATTCGCAAGTTTAATTCGCAATTCATTT 360

OY 361 CAGAAACAGGTTTAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 420
DB 361 CAGAAACAGGTTTAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 420
OY 421 GCTATCATGTTTGTCTTACTTAATTT 447
DB 421 ATCATCTTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 447

RESULT 3
US-10-793-626-3743

; Sequence 3743, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3743
; LENGTH: 3240
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3743

Query Match 40.7%; Score 182; DB 8; Length 3240;
Best Local Similarity 65.5%; Pred. No. 1.1e-40;
Matches 266; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

OY 1 ATGTTGGATCGATTTATACATCGACATTAATTTAGGTGATCTACGATTAATTTT 60
DB 2834 ATGATAGGTTCCGAATTTATTTCTCTTATTCGAGGTTGTCGATCTCAAGTTGATTTT 2893
OY 61 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGTACCGGATTTTATGAGCTTGTG 120
DB 2894 GCTGAGAAATTTGGATTAATCCAGAGGTTAGTGTTCCAGGTTTATTTAGCTTTGATTT 2953
OY 121 TTTAATCAGCGGCTTTATTTTACTTGTGTTGCTAGTAGAGCTTCTCATTTAGTTATC 180
DB 2954 TTTGATCAACCGATCATGTTGTTATTCAGATTAATCATTTAGTCTTAACTTAATTTATC 3013
OY 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
DB 3014 GTAGCAACGGTATTAAGTAGGTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3073
OY 241 CTGATTAACAGGATGCTCTTAAATTCGCGTTTGAATTTTCTATACCGGATTTGATTAATTT 300
DB 3074 ATACTGACGGGATGATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3133
OY 301 GAAATCGAGAAATTTTCGAGGATTCGGCATTCGCGATTCGCAAGTTTAATTCGCAATTCATTT 360
DB 3134 GAAATCGAGAAATTTTCGAGGATTCGGCATTCGCGATTCGCAAGTTTAATTCGCAATTCATTT 3193
OY 361 CAGAAACAGGTTTAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 406
DB 3194 CAGAAACAGGTTTAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 3239

RESULT 4
US-10-793-626-4197/c

; Sequence 4197, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US

CURRENT APPLICATION NUMBER: US/10/793,626
CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4197
LENGTH: 3029
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-4197

Query Match 10.2%; Score 45.6; DB 8; Length 3029;
Best Local Similarity 62.1%; Pred. No. 0.011;
Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 332 TCGTCCAGTTTAATGCAATACCAATGAGAAACAGTTTAACCACTTACGTCGGA 391
DB 3023 TCATTCCTGATATATGGAATCAATCAAAACAGTGTAGTCACTTACCTTCTA 2964
QY 392 GCAGCTGCTATGAGCGGAGGACCTTGTATCATGTTTGTACTACTAATT 447
DB 2963 CACCAATGTTTAATCAATGATATATATATATATATATATATATATATAT 2908

RESULT 5
US-11-181-234-1
Sequence 1, Application US/1181234
Publication No. US20060021075A1
GENERAL INFORMATION:
APPLICANT: WANG, CHYUNG-RU
TITLE OF INVENTION: GROUP 1 CD1 TRANSGENIC MICE AND THEIR
FILE REFERENCE: 21117.0001U2
CURRENT APPLICATION NUMBER: US/11/181,234
CURRENT FILING DATE: 2005-07-14
PRIOR APPLICATION NUMBER: 60/588,192
PRIOR FILING DATE: 2004-07-15
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 168753
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: note =
US-11-181-234-1

Query Match 8.5%; Score 37.8; DB 12; Length 168753;
Best Local Similarity 49.3%; Pred. No. 6.9; Indels 102; Gaps 0;
Matches 99; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 15 TTTATACATGCACTAATTTAGTGTACTACTAGTTTAATTTTGGCGAAAAACAG 74
DB 91496 TTATATACATTTGCTGCTGTTAGTTCATTTGCTAGTTTCAAGTAAACAAACA 91555
QY 75 GATCGTCCGCGAGCACTGTTGTACCGGATATTTAGACCTTGTGTTTAATCAGCCGGT 134
DB 91556 CCGCGACTTTTGAATGCTGAGATATGTGTGATTCCTCTCGTGGAGGAGCTGGC 91615
QY 135 CTTTATTTACTGTTTGTCTAGTACCTTGCCTCACTATATGCTGAAATACGTTT 194
DB 91616 ACTTCTTTTCTTTTGAATTTGTGCAATTTGATGAGGTTTATATATATATATAT 91675
QY 195 ATCCAAATTTATATATTTGTA 215
DB 91676 TTTTATATATATATATATATA 91696

RESULT 6
US-09-925-065A-307190
Sequence 307190, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 307190
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-307190

Query Match 7.5%; Score 33.6; DB 6; Length 389;
Best Local Similarity 52.1%; Pred. No. 12;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 65 AAAAAACAGGATCGTCCGCGAGCACTTGTGTAACCGGATATTAGACCTTGTTT 124
DB 70 AAAATCTGTTATTAAGTAGTACCACTTGATCAATTAATCTTAGCTTCTTGAGATA 129
QY 125 ATGACCCGCTCTTATTTTACTGTTTCTAGTGTGCTCACTTATGTTATGTA 184
DB 130 ATTCTCTGAGCTTCTGATTAACCTTGCTGTTACCTTGACCTTATGTTATGAGA 189
QY 185 AATACGTTTATCAATTTATGA 208
DB 190 CAGCTTCTTATGAATTTTATGA 213

RESULT 7
US-09-925-065A-297021/c
Sequence 297021, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 297021
LENGTH: 498
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-297021/c

US-09-925-065A-297021

Query Match 7.5%; Score 33.4; DB 6; Length 498;
Best Local Similarity 52.5%; Pred. No. 14;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 92 TTGTTGACGGGATTTAGAGCTTGTGTTAATCAGCGGCTTTTACTTGTGTT 151
DB 308 TTGTCCTCAATACACTAGACCTTATTTCTCTATCTAGATATTTCTTCTTCTT 249
QY 152 TGCTAGAGCTTGTCTCACTTATGCGAAATACGGTTTATCCAAATTTATGATTT 211
DB 248 TCTTTTAAATTTAGTTTCTTCTAGCTTCAAGAGTAGCTTTATCCAAATTTATAT 189
QY 212 TGTAACGACGAGAAATTT 230
DB 188 TGTAATTTTAAATTAAT 170

RESULT 8

US-09-925-065A-780420/C

; Sequence 780420, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925.065A

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 780420

; LENGTH: 528

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-780420

Query Match 7.5%; Score 33.4; DB 6; Length 528;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 61 GCGGAAAAACAGGATCGTCCGACAGACTTGTGACCGGATTTAGAGCTTGTG 120
DB 500 GAGGAGAGAGAGAGAGCTCAAGCAGGACAGCTGCTCGGACCTCCCTTTTGA 441
QY 121 TTTATGACCGCGCTTATTTTACTTGTGTTGCTAGAGAGCTTCTCACTATGATTC 180
DB 440 TTTAAGCAGACATTTATTTTGTGCTCGATTAATTTGATTTTGGATTTAGTAAATC 381
QY 181 GTGAATACGGTTTATCCAAATTTTATGATTTTGTACGAGCAGAAATTCGCTCCATG 240
DB 380 CAAAACATTTTTCGCAAAATGACAAAGTGGTTAGAAAATCTGAAATTCCTACAGTG 321
QY 241 CTGATTA 247
DB 320 AAGATGA 314

RESULT 9

US-09-925-065A-804269/C

; Sequence 804269, Application US/09925065A
; Publication No. US20040181048A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925.065A

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 804269

; LENGTH: 530

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-804269

Query Match 7.5%; Score 33.4; DB 6; Length 530;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 61 GCGGAAAAACAGGATCGTCCGACAGACTTGTGACCGGATTTAGAGCTTGTG 120
DB 500 GAGGAGAGAGAGAGAGCTCAAGCAGGACAGCTGCTCGGACCTCCCTTTTGA 441
QY 121 TTTATGACCGCGCTTATTTTACTTGTGTTGCTAGAGAGCTTCTCACTATGATTC 180
DB 440 TTTAAGCAGACATTTATTTTGTGCTCGATTAATTTGATTTTGGATTTAGTAAATC 381
QY 181 GTGAATACGGTTTATCCAAATTTTATGATTTTGTACGAGCAGAAATTCGCTCCATG 240
DB 380 CAAAACATTTTTCGCAAAATGACAAAGTGGTTAGAAAATCTGAAATTCCTACAGTG 321
QY 241 CTGATTA 247
DB 320 AAGATGA 314

RESULT 10

US-09-925-065A-547674

; Sequence 547674, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925.065A

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 547674

; LENGTH: 569

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-547674

Query Match 7.5%; Score 33.4; DB 6; Length 569;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

OY 116 TTGCTTTAATCAGCCGCTTTATTTACTTTGCTGAGCTGCTGCACTTANG 175
|||
DB 305 TTGACTGAGACTTTTGATGTTACTTTCCAAATTTCTAGTTGCTCCACACAAAT 364
176 TTATCGTAATAGCGTTTATCCAAATTTATGATTTTGAGCGAGCAGAAATTCGCTG 235
DB 365 TTGACAAAGACTTTCTTTTACGACATTTATTTTAAATATCTTTCAAAACATTTCACT 424
OY 236 CCATGCTGATTAACAGGATTCGCTTAAATGCGTTGATTTCTATACCGATTTGAC 295
DB 425 TAGTTTCAAAACATTTCACTTCTTCAAACTTTGTAAGAGATTAACACTTTTGTG 484
OY 296 CATTGGA 302
|||
DB 485 CATTTAA 491

RESULT 11

US-09-925-065A-547675
; Sequence 547675, Application US/09925065A
; Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 547675
LENGTH: 569
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-547675

Query Match 7.5%; Score 33.4; DB 6; Length 569;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

OY 116 TTGCTTTAATCAGCCGCTTTATTTACTTTGCTGAGCTGCTGCACTTANG 175
|||
DB 305 TTGACTGAGACTTTTGATGTTACTTTCCAAATTTCTAGTTGCTCCACACAAAT 364
OY 176 TTATCGTAATAGCGTTTATCCAAATTTATGATTTTGAGCGAGCAGAAATTCGCTG 235
DB 365 TTGACAAAGACTTTCTTTTACGACATTTATTTTAAATATCTTTCAAAACATTTCACT 424
OY 236 CCATGCTGATTAACAGGATTCGCTTAAATGCGTTGATTTCTATACCGATTTGAC 295
DB 425 TAGTTTCAAAACATTTCACTTCTTCAAACTTTGTAAGAGATTAACACTTTTGTG 484
OY 296 CATTGGA 302
|||
DB 485 CATTTAA 491

RESULT 12

US-09-925-065A-418523
; Sequence 418523, Application US/09925065A
; Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 418523
LENGTH: 611
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-418523

Query Match 7.5%; Score 33.4; DB 6; Length 611;
Best Local Similarity 52.5%; Pred. No. 15;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 92 TTGTTGATCCGGATATTAGACTTGTGTTATCAGCCGCTTTATTTACTTGT 151
|||
DB 194 TTGCTCAATCAATCACTAGACTTATTTCTTCAATGATGATTTTCTTTCTT 253
OY 152 TGCTAGTGAAGCTTGTCTACTATGTTATCGGAATAGCGTTTATCCAAATTTATGATT 211
DB 254 TCTTTTAAATTAATGTTTCTTAGGTTTCAATGAGGTTTATCCAAATTTATTAATT 313
OY 212 TGTAAGAGCAGCAAAATT 230
DB 314 TGTATGATTTAGAAATAT 332

RESULT 13

US-09-925-065A-418524
; Sequence 418524, Application US/09925065A
; Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 418524
LENGTH: 611
TYPE: DNA
ORGANISM: Homo sapiens

US-09-925-065A-418524

Query Match 7.5%: Score 33.4; DB 6; Length 611;
Best Local Similarity 52.5%: Pred. No. 15;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 92 TTGTTGACCGGATATTAGACTTGTGTTAATCAGCCGCTCTTATTACTTGT 151
Db 194 TTGTCATCAATACACTAGACCTTATTTCTCTATCTAAGTATTTTCTTCTTCT 253
Qy 152 TGTAGTACGCTTGTCTACTATATGTAATCGGAATACGTTTATCCAAATTATGATT 211
Db 254 TCTTTTAATTAGTTTCTTAGGTTTCATGAGTAGCTTTATCCAAATTATTAAT 313
Qy 212 TGTACGACGCGAATAATT 230
Db 314 TGTATGTAATTTAGAAATAAT 332

RESULT 14

US-09-925-065A-692556

Sequence 692556, Application US/09925065A
Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single

FILE REFERENCE: 108827.135

CURRENT APPLICATION NUMBER: US/09/925.065A

PRIOR FILING DATE: 2001-08-08

PRIOR FILING DATE: 2000-10-24

PRIOR FILING DATE: 2000-11-20

PRIOR FILING DATE: 2000-11-20

PRIOR FILING DATE: 2000-11-30

PRIOR FILING DATE: 2001-01-16

PRIOR FILING DATE: 2001-05-09

NUMBER OF SEQ ID NOS: 957086

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 692556

LENGTH: 1524

TYPE: DNA

ORGANISM: Homo sapiens

US-09-925-065A-692556

Query Match 7.5%: Score 33.4; DB 6; Length 1524;
Best Local Similarity 52.5%: Pred. No. 21;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 92 TTGTTGACCGGATATTAGACTTGTGTTAATCAGCCGCTCTTATTACTTGT 151
Db 221 TTGTCATCAATACACTAGACCTTATTTCTCTATCTAAGTATTTTCTTCTTCT 280
Qy 152 TGTAGTACGCTTGTCTACTATATGTAATCGGAATACGTTTATCCAAATTATGATT 211
Db 281 TCTTTTAATTAGTTTCTTAGGTTTCATGAGTAGCTTTATCCAAATTATTAAT 340
Qy 212 TGTACGACGCGAATAATT 230
Db 341 TGTATGTAATTTAGAAATAAT 359

RESULT 15

US-09-925-065A-692557

Sequence 692557, Application US/09925065A
Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single

FILE REFERENCE: 108827.135

CURRENT APPLICATION NUMBER: US/09/925.065A

PRIOR FILING DATE: 2001-08-08

PRIOR FILING DATE: 2000-10-24

PRIOR FILING DATE: 2000-11-20

PRIOR FILING DATE: 2000-11-30

PRIOR FILING DATE: 2001-01-16

PRIOR FILING DATE: 2001-05-09

NUMBER OF SEQ ID NOS: 957086

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 692557

LENGTH: 1524

TYPE: DNA

ORGANISM: Homo sapiens

US-09-925-065A-692557

FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925.065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 692557
LENGTH: 1524
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-692557

Query Match 7.5%: Score 33.4; DB 6; Length 1524;
Best Local Similarity 52.5%: Pred. No. 21;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 92 TTGTTGACCGGATATTAGACTTGTGTTAATCAGCCGCTCTTATTACTTGT 151
Db 221 TTGTCATCAATACACTAGACCTTATTTCTCTATCTAAGTATTTTCTTCTTCT 280
Qy 152 TGTAGTACGCTTGTCTACTATATGTAATCGGAATACGTTTATCCAAATTATGATT 211
Db 281 TCTTTTAATTAGTTTCTTAGGTTTCATGAGTAGCTTTATCCAAATTATTAAT 340
Qy 212 TGTACGACGCGAATAATT 230
Db 341 TGTATGTAATTTAGAAATAAT 359

Search completed: February 27, 2006, 11:38:28
Job time : 270.374 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 105.186 Seconds
(without alignments)
7553.984 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatagattata.....tcttttcttactactaatt 447

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/1_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/H_COMB.seq:*
6: /cgn2_6/prodata/1/ina/CCTUS_COMB.seq:*
7: /cgn2_6/prodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/prodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/prodata/1/ina/backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	189.4	42.4	453	US-09-710-279-2323	Sequence 2323, Ap
2	189.4	42.4	459	US-09-134-001C-1666	Sequence 1666, Ap
3	182	40.7	3240	US-09-710-279-3743	Sequence 3743, Ap
4	52.8	11.8	198	US-09-134-001C-1489	Sequence 1489, Ap
5	45.6	10.2	3029	US-09-710-279-4197	Sequence 4197, Ap
6	35.2	7.9	4078	US-09-928-847B-43	Sequence 43, Appl
7	34.4	7.7	534	US-09-593-995-6	Sequence 6, Appl
8	34.4	7.7	534	US-10-139-667-6	Sequence 6, Appl
9	34.4	7.7	51259	US-08-781-891-209	Sequence 209, App
10	34.4	7.7	51259	US-09-618-166-209	Sequence 209, App
11	34.2	7.7	1266	US-09-949-016-16928	Sequence 16928, A
12	34.2	7.7	1266	US-08-961-527-137	Sequence 137, App
13	34.2	7.7	260247	US-09-949-016-14338	Sequence 14338, A
14	33.6	7.5	96109	US-09-596-002-35	Sequence 35, Appl
15	33.2	7.4	501	US-09-248-796A-434	Sequence 434, App
16	33.2	7.4	524032	US-09-949-016-16928	Sequence 16928, A
17	33.2	7.4	524032	US-09-949-016-16928	Sequence 16928, A
18	33.2	7.4	524032	US-09-949-016-16928	Sequence 16928, A
19	33.2	7.4	524032	US-09-949-016-16928	Sequence 16928, A
20	33.2	7.4	529885	US-09-949-016-14340	Sequence 14340, A
21	33.2	7.4	529885	US-09-949-016-14341	Sequence 14341, A
22	33.2	7.4	529885	US-09-949-016-14342	Sequence 14342, A
23	33.2	7.4	529885	US-09-949-016-14343	Sequence 14343, A
24	33.2	7.4	529885	US-09-949-016-14344	Sequence 14344, A

25	33.2	7.4	529885	3	US-09-949-016-14345	Sequence 14345, A
26	33.2	7.4	529885	3	US-09-949-016-14346	Sequence 14346, A
27	33.2	7.4	529885	3	US-09-949-016-14347	Sequence 14347, A
28	33	7.4	2468	3	US-09-559-23	Sequence 23, Appl
29	33	7.4	6004	3	US-08-961-527-27	Sequence 27, Appl
30	32.8	7.3	474	3	US-09-621-976-1803	Sequence 1803, A
31	32.8	7.3	85665	3	US-09-949-016-17348	Sequence 17348, A
32	32.6	7.3	1106	3	US-09-107-433-1435	Sequence 1435, Ap
33	32.6	7.3	1347	3	US-09-583-110-1828	Sequence 1828, Ap
34	32.4	7.2	601	3	US-09-949-016-66363	Sequence 66363, A
35	32.4	7.2	1929	3	US-09-248-796A-772	Sequence 772, App
36	32.4	7.2	20441	3	US-09-949-016-14476	Sequence 14476, A
37	32.4	7.2	24645	3	US-09-949-016-14062	Sequence 14062, A
38	32.4	7.2	422592	3	US-09-949-016-14182	Sequence 14182, A
39	32.2	7.2	139552	3	US-09-949-016-15300	Sequence 15300, A
40	32	7.2	601	3	US-09-949-016-47889	Sequence 47889, A
41	32	7.2	601	3	US-09-949-016-47890	Sequence 47890, A
42	32	7.2	601	3	US-09-949-016-94477	Sequence 94477, A
43	32	7.2	1220	2	US-08-940-767-3	Sequence 3, Appl
44	32	7.2	1220	3	US-08-941-155-3	Sequence 3, Appl
45	32	7.2	1220	3	US-09-344-666-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-710-279-2323
; Sequence 2323, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS48005
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2323
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-2323

Query Match	42.4%	Score 189.4	DB 3	Length 453
Best Local Similarity	64.0%	Pred. No. 2.8e-51		
Matches 286	Conservative 0	Mismatches 161	Indels 0	Gaps 0
1	ATGTCGATCAGATTATATACATTCGACTAATTTAGTGTACTACTAGTTAATTTT	60		
1	ATGTCGATCAGATTATATATATTTCTTATTCGTAAGTGTACTAGTTAATTTT	60		
1	ATGTCGATCAGATTATATATATTTCTTATTCGTAAGTGTACTAGTTAATTTT	60		
61	GGCGAAAAACAGAGATCGTCCGCGAGACTTTGTACCGGATTTTATAGACTTGG	120		
61	GGCGAAAAATTTGGATTAATCCAGCAGCGGTAGTGTTCAGGTTAATTTAGTTGAT	120		
121	TTTAAATCAGCGGTCTTTATTTTACTGTTTTCGTAAGTGTCTGCTCAATATGTTAT	180		
121	TTTAAATCAGCGGTCTTTATTTTACTGTTTTCGTAAGTGTCTGCTCAATATGTTAT	180		
121	TTTAAATCAGCGGTCTTTATTTTACTGTTTTCGTAAGTGTCTGCTCAATATGTTAT	180		
181	GTCGAATACGGATTATATCAAAATTTATGTCGAGCGAGAAATTCGTCGCATG	240		
181	GTCGAATACGGATTATATCAAAATTTATGTCGAGCGAGAAATTCGTCGCATG	240		
181	GTCGAATACGGATTATATCAAAATTTATGTCGAGCGAGAAATTCGTCGCATG	240		
241	CTGATACAGAGGATCGTCTTAAATTCGCTTGTATCCGATTTGACATTT	300		
241	ATACCTACGAGATGTCATTAATTTATTTGATCTCTTGTACCATTCACCATTT	300		

Qy	301	GAATTCGAGAAATTCGAGGAATCGGACATCATCGTCCAGGTTTAATGCCAATCCATT	360
Qy	301	GAATTCGAGAAATTCGAGGAATCGGACATCATCGTCCAGGTTTAATGCCAATCCATT	360
Db	301	GAATTCGTTGAAGTTTCAGTATAGGTGTGTGTCATTCCTGGTAATTAATTCGCAATCAATT	360
Qy	361	CAGAAACAGGTTTAAACCATTAAGTTGCGAAGCAGCTGCTATTGAGCGGACGACCTTT	420
Db	361	CAAAACAGGTTGATGTCATTACACTTCTACACAAATGTTATTAACTATGTTACATAT	420
Qy	421	GCTATCATGTTTGTTTACTACTTAATT	447
Db	421	ATCATCTTATTTTATATATAGTTTAA	447

```

RESULT 2
US-09-134-001C-1666
: Sequence 1666, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1666
: LENGTH: 459
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-1666

```

Query Match	42.4%	Score 189.4	DB 3	Length 459
Best Local Similarity	64.0%	Pred. No. 2.8e-51		
Matches 286; Conservative	0	Mismatches 161	Indels 0	Gaps 0

Oy	1	ATGTCGGATTCAGATTTTATACATCCGACATMAATTTTAAAGTGTA	60
Db	7	ATGATAGGTTGCAAAATTATATTCTTCCTATTCCGAGGTGCTGACTCAGTTGATATTT	66
Oy	61	GGCGAAAAAACAGGGATCGTCCCGCAGACCTGTTGTACCGGGATTTTAGACCTGTG	120
Db	67	GCTGGAAATTTGGGATTTAATCCACAGGGTTAGTCGTTCCAGGTTATTTACCTTTGAT	126
Oy	121	TTTATATCAGCCGGTCTTTATTTTACTGTGTTTGGTAGTAGCGTGTCACTATTGTATC	180
Db	127	TTTGATCAACCGATCATGTGTTATCAGATTTAATCATTTAGTTGCTTAATTTTATTC	186
Oy	181	GTGAATAACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGTGCGATG	240
Db	187	GTAACCAACGGTATTAGTTAAGTAGGGTTATTTTATATGTAGAGAAATAATCCGTGCGATG	246
Oy	241	CTGATTAACAGGGATCGTCTTAATAAATCCGGTTTGATTTTCTATACCCGATTTGACATTT	300
Db	247	ATACGACGGGAAATGATTAATTTAATTTATATTTGATCTCTTGTACCCATTTGACCCATTT	306
Oy	301	GAATTCGACGAATTTTCGAGGAATCGGCATCATCTGTCGAGTTTAATTCACATACATT	360
Db	307	GAATTCGTTGAAGTTTCAGGTATAGGTGTTCATCTCCGTGATTTATTCGAAATCAATTT	366
Oy	361	CAGAAACAAGTTTAACCATTAACGTTCCGAGACGCTGCTATTTGACGCGAGGACCTTT	420
Db	367	CAAAAACAAGGTGTGATCATTTACACTTCTACACAATGTTATTAACATGATTTACATAT	426
Oy	421	GCTATCATGTTGTTTACTACTTAATT	447
Db	427	ATCATCTTATTTTATATAGTTTATTT	453

```

US-09-710-279-3743
: Sequence 3743, Application US/09710279
: Patent No. 6703492
: GENERAL INFORMATION:
: APPLICANT: KIMBERLY, WILLIAM JOHN
: TITLE OF INVENTION: STRAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
: FILE REFERENCE: PU348005
: CURRENT APPLICATION NUMBER: US/09/710,279
: CURRENT FILING DATE: 2000-11-09
: PRIOR APPLICATION NUMBER: 60/164,258
: PRIOR FILING DATE: 1999-11-09
: NUMBER OF SEQ. ID NOS: 4472
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3743
: LENGTH: 3240
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: synthetic
: OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3743

```

Query March	40.7%	Score 182	DB 3	length 3240
Best Local Similarity	65.5%	Pred. No.	17e-48	
Matches 266	Conservative	0	Mismatches 140	Indels 0
			Gaps	0

Oy	1	ATGTTGGATCAGATTATATACATCGCACTAATTTAGTGTACATCAACGATTAATTTT	60
Db	2834	ATGATAGGTCAGAAATATATTTCTCTATTGCTAGGTGTGACTACAGTTATATTT	2893
Oy	61	GCGGAAAAAACAGGAGATCGTGC CGGCAAGACTTGTGTACCGGATATTTAGACTGTG	120
Db	2894	GCTAGAAATTTGGGATTAATCCAGCAGGGTTACTCGTTCCAGGTTATTAGCTTGTATT	2953
Oy	121	TTTATACGCCGCTTTATTTTACTTGTTTGCTAGTACCTTGCTCACTTAGTTATC	180
Db	2954	TTTATATCAACCGATCATGTTGTTATCAGTATTAATCAATTAGTCTTAACTTAATTTATC	3013
Oy	181	GTGAATAACGGTTATATCCAAATTTATGATTTTGTACGAGCAGAAAAATTCGCTGCATG	240
Db	3014	GTAAGCAACGGTATATAGTAAGTGGATTATTTATATGTAGAAAGAAAATTCGCTGCATG	3073
Oy	241	CTGATTAACAGGAGTCGTCTTAAAAATGCGCTTGATTTTCTAATCCGATTTGACATTT	300
Db	3074	ATACTGACGGGAATGGTGATTAATTTATATTGATCTCTTGATCCCATGACCCCATTT	3133
Oy	301	GAAATCGCAGATTTCCAGAAATGCGCATCATGTCGCAAGTTTAATTTGCCAATACCAT	360
Db	3134	GAAATGGTGAAGTTTCAGGATATGGGTGTCTACTTCGTGATATTATTTGCGCAATACATTT	3193
Oy	361	CAGAAACAGGTTTAACCATTCGTTGCGGAAGCAGCGCTGATTTGA	406
Db	3194	CAAAAACAGGTGTAGTCATTATACCTTTCTTCAACAATGTTATTTATA	3239

```

RESULT 4
US-09-134-001C-1489/c
; Sequence 1489, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR APPLICATION NUMBER: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1489
; LENGTH: 198

```

TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1489

Query Match 11.8%; Score 52.8; DB 3; Length 198;
Best Local Similarity 69.2%; Pred. No. 3.6e-07;
Matches 72; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGTCGATCAGATTATACATCGCACTAATTAGGTACTACTAGTTAATTTT 60
DB 104 ATGATAGTTGAGATTATATTTCTTATTCGAGGTGTCGACTGATGATATT 45
QY 61 GCGGAAAAACAGGATGTCGCCGACGACTGTTTGAACCGG 104
DB 44 GCTGAGAAATTTGGATTATTCACGAGGTTAGTGTTCAGG 1

RESULT 5
US-09-710-279-4197/c

Sequence 4197, Application US/09710279
Patent No. 6703492

GENERAL INFORMATION:

APPLICANT: KIMMERLY, WILLIAM JOHN

TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

FILE REFERENCE: P0348005

CURRENT APPLICATION NUMBER: US/09/710,279

CURRENT FILING DATE: 2000-11-09

PRIOR APPLICATION NUMBER: 60/164,258

NUMBER OF SEQ ID NOS: 4472

SOFTWARE: Patent Ver. 2.1

SEQ ID NO 4197

LENGTH: 3029

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

US-09-710-279-4197

Query Match 10.2%; Score 45.6; DB 3; Length 3029;
Best Local Similarity 62.1%; Pred. No. 0.00027;
Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 332 TCGTCCAGGTTTAAATGCAATACCATTCAGAAACAGGTTAACATTACGTTGAA 391
DB 3023 TCAATCTCGTATATATGGAATCAATCAAAAGGTGATGATTAACACTTTCTA 2964

QY 392 GCACCTGCTATTAGCGGAGCAGCCTTGTCTATCATGTTTGTACTTAATT 447
DB 2963 CAACATGTTATTAACATGATATATATCATCTTATTTTATATAGTTTATT 2308

RESULT 6
US-09-928-847B-43

Sequence 43, Application US/09928847B

Patent No. 680896

GENERAL INFORMATION:

APPLICANT: No. 680896ozymes A/S

APPLICANT: Jorgensen, Steen T

APPLICANT: Rasmussen, Michael D

APPLICANT: Andersen, Jens Tonne

APPLICANT: Olsen, Carsten

TITLE OF INVENTION: Multiple Insertion of Genes

FILE REFERENCE: 10022.204-US

CURRENT APPLICATION NUMBER: US/09/928,847B

CURRENT FILING DATE: 2002-05-06

NUMBER OF SEQ ID NOS: 50

SOFTWARE: Patent version 3.1

SEQ ID NO 43

LENGTH: 4078

TYPE: DNA

ORGANISM: Bacillus licheniformis

US-09-928-847B-43

Query Match 7.9%; Score 35.2; DB 3; Length 4078;
Best Local Similarity 52.8%; Pred. No. 0.72; 68; Indels 0; Gaps 0;
Matches 76; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 203 TTATGATTTGTAGCAGCAGCAAAATTCGTGCGCATGCTGATTAACAGGATGCTCTAA 262
DB 2799 TTACATTTTAACTGATGCTCTTTCGCCGCTATAGCTGCCGAGAAAGATCGCTTCC 2858
QY 263 AAATCGCTTGATTTCTATACCCGATTTGACATTTGAAATGCGAAATTTGAGGAA 322
DB 2859 AAATCGCTGCGCCCTCCGCAAGAGGTTGAAGGCTTGAAGGAGGATTCATCAATCA 2918
QY 323 TCGCATCATGTCGCCAGGTTAA 346
DB 2919 TTCAGTGTGATGAACCTGCTTGA 2942

RESULT 7

US-09-593-995-6/c

Sequence 6, Application US/09593995

Patent No. 6406888

GENERAL INFORMATION:

APPLICANT: Conklin, Darrell C.

TITLE OF INVENTION: HELICAL CYTOKINE ZALPHA33

FILE REFERENCE: 99-38

CURRENT APPLICATION NUMBER: US/09/593,995

CURRENT FILING DATE: 2000-06-14

PRIOR APPLICATION NUMBER: 60/139,121

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 6

LENGTH: 534

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: degenerate sequence

NAME/KEY: misc.feature

LOCATION: (1)...(534)

OTHER INFORMATION: n = A,T,C or G

US-09-593-995-6

Query Match 7.7%; Score 34.4; DB 3; Length 534;
Best Local Similarity 28.3%; Pred. No. 0.52;
Matches 73; Conservative 44; Mismatches 141; Indels 0; Gaps 0;

QY 106 TATTTAGACTGTGTTTAATACGCGGCTTTATTTACTGTTGCTAGTACTTG 165
DB 510 DATRTCCADATNARRTNCKNARRTGRTCCATDATTGTTNARYTGTGTTTGTG 451

QY 166 CTCACCTATGTTATGCTGAATATACGTTTATCAATTTATGATTTTGTACGACGAGA 225
DB 450 YTTNARYTTTTRTNACTTCNGCDATTCNCKNKTCYTNSNNGCAANCKXGANG 391

QY 226 AAATTCGCTGCATGCTGATAACAGGATGCTCTTAAATGCGCTTCAATTTCTATAC 285
DB 390 NCONGNCNCRTRCTRTTYTNSMNCRCRYCYTCNACRTANAGDATTNARYTGYTCNAC 331

QY 286 CCGATTGTACATTTGAAATGCGAGATTTGAGAGATTCGAGATCATGCTGCAGGTTTA 345
DB 330 NEGDAATNGRTCCATNCCNCRCAATTTTCRTTCAATTTTCTANACNARNCKNARYTT 271

QY 346 ATTGCAATACCATTCAG 363
DB 270 NCRRAANRNACNSMVAR 253

RESULT 8

US-10-139-667-6/c

Sequence 6, Application US/10139667

Patent No. 6761882
GENERAL INFORMATION:
APPLICANT: Conklin, Darrell C.
APPLICANT: Gao, Zeren
TITLE OF INVENTION: HELICAL CYTOKINE ZALPHA33
FILE REFERENCE: 99-38
CURRENT APPLICATION NUMBER: US/10/139,667
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/09/593,995
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: 60/139,121
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 6
LENGTH: 534
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: degenerate sequence
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(534)
OTHER INFORMATION: n = A,T,C or G
US-10-139-667-6

Query Match 7.7%; Score 34.4; DB 3; Length 534;
Best Local Similarity 28.3%; Pred. No. 0.52;
Matches 73; Conservative 44; Mismatches 141; Indels 0; Gaps 0;

QY 106 TATTAGACTTGTGTTATATCAGCCGCTTTATTTACTGTTGCTAGTGAAGCTTG 165
DB 510 DATRCCCADATNARTTNCNARTRTCCATDATTGTTNARTGTGTTTGTG 451
QY 166 CTCACCTATGTTATGCTGAATAATCGCTTTATCCAAATTTATGATTTTGAAGCAGCA 225
DB 450 YTTNARTTTRTTNACATGCGDATATCNCNCKYCYCNCSWNGCRANCKNGGNG 391
QY 226 AAATTCGCTGCATGCTGATACAGGATCGCTTAAATCGCTTGAATTTCTATAC 285
DB 390 NCCGNCNCRTRCTRTTNTNSMNCRCRTCYTCNACRTANAGDATNARTGTTCNAC 331
QY 286 CCGATTGTACATTTGAATCGCAGAAATTCGAGAAATCGCATGCTGCAGGTTTA 345
DB 330 NGDATTNGRCCATNCNCRCGATTTTCRTTRCATTTTCRTANACNCKNARTT 271
QY 346 ATTGCAATACCATTCAG 363
DB 270 NCKRANARNACNSWNR 253

RESULT 9
US-08-781-891-209
Sequence 209, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Yu, Ying-Hui
APPLICANT: Yu, Chang-Hu
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-209

Query Match 7.7%; Score 34.4; DB 3; Length 51259;
Best Local Similarity 51.3%; Pred. No. 4.1;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 105 ATATTGACCTTGTTATATCAGCCGCTTTATTTACTGTTGCTAGTGAAGCTT 164
DB 25726 ATATTGACCTTGTTATATCAGCCGCTTTATTTACTGTTGCTAGTGAAGCTT 25785
QY 165 GCTCCTATGTTATGCTGAATAATCGCTTTATCCAAATTTATGATTTTGAAGCAGCAG 224
DB 25786 ATGCGCTGTTATGCTGCTGTTAGGTTATCATCTGTTGATTTTTCAGAGAA 25845
QY 225 AAAATTCGCTGCATGCTGATACAGGATGCTT 260
DB 25846 CCACTCTGATTTGTTGATTTCTTGTATAGTCT 25881

RESULT 10
US-09-618-166-209
Sequence 209, Application US/09618166
Patent No. 6583112
GENERAL INFORMATION:
APPLICANT: Yu, Ying-Hui
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/618,166
FILING DATE: 17-Jul-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052.419C1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-09-618-166-209

Query Match 7.7%; Score 34.4; DB 3; Length 51259;
Best Local Similarity 51.3%; Pred. No. 4.1;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

105 ATATTGAGCTGTGTTTAATCAAGCCGCTTATTTACTGTTTGTAGTACGCTT 164
25726 ATATTGAGCTGTGTTTAATCAAGCCGCTTATTTACTGTTTGTAGTACGCTT 25785

165 GCTCACTTATGTTATCGTAATACGCTTATTCGAATTTGTTGTAAGGAGCGAG 224

25786 ATGCCCTCTGTTACTCTGCTAGGCTTATCTATCTGTTGATTTTTCAGAGAA 25845

225 AAAATGCGTCGCAATGCTGATTAACAGGATGCTCT 260
25846 CCAGCTCCTAGTTTGTGATTTCTTGTATAGTTCT 25881

RESULT 11
US-09-949-016-55321
Sequence 55321, Application US/09949016

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 55321
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-55321

Query Match 7.7%; Score 34.2; DB 3; Length 601;
Best Local Similarity 49.2%; Pred. No. 0.64;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

83 CGGAGAGCTGTGTTGACCGGATTTTAGACTGTGTTTATACGCGGCTTATTT 142

106 CAGCAGCATTATTAAGCAAGGATTAATTTCCCATGCTCATACACGATTTTATTT 165

143 TACTGTTTGTAGTACGCTTGTCTCATTTATGTTAATCGTAATACGGTTATTCGAAT 202

166 ACTGATCTTAGTTGTTGTTTGTATTTCTGTGATCTTTTCAAACTCACTATAAGTT 225

203 TTATGATTTTGAAGCAGCAAAATTCGCTGCCATGCTGATTAACAGGATGCTCTAA 262

226 TTAAAGATATTTAGCATATGTCATAACATTTTCTACATGAAAAATCAATGTCATCATCA 285

263 AAA 265

286 AAA 288

RESULT 12

US-08-961-527-137/c
Sequence 137, Application US/08961527

PATENT NO. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA

ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527

FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:

LENGTH: 1266 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-137

Query Match 7.7%; Score 34.2; DB 3; Length 1266;
Best Local Similarity 52.4%; Pred. No. 2.5; 68; Indels 0; Gaps 0;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

121 TTATATGACCGGCTTATTTTACTGTTTGTCTAGTACGCTTCTCATTTATGTTATC 180

12235 TTGACGTACTCTGCTGATATCGGTGTTGTGACGCGTAATTTGTTACCTTACGGTC 12176

181 GTGAATACGCTTTATTCGAATTTTATGATTTTGAAGCAGCAGAAAAATTCGCTGCATG 240

12175 AATACAAAGCTTATACCAATTTGATGCTGAGTCTTAACCTTTGATTTG 12116

241 CTGATTAACAGGATGCTCTTAA 263

12115 GTGTAAGCTTGAATGCTCCAGAA 12093

RESULT 13

US-09-949-016-13358/c
Sequence 13358, Application US/09949016

PATENT NO. 6812339
GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13358
LENGTH: 260247
TYPE: DNA
ORGANISM: Human
US-09-949-016-13358

Query Match 7.7%; Score 34.2; DB 3; Length 260247;
Best Local Similarity 49.2%; Pred. No. 9.9; Mismatches 93; Indels 0; Gaps 0;
Matches 90; Conservative 0;

QY 83 CGGAGAGACTGTTGTACCGGAGATTTAGACCTGTGTTAATCAGCGGCTTTATT 142
DB 112790 CAGCACCATTTATGAAACAGGGAATTAATTTCCCATTCCTCATACACGATTTTATT 112731
QY 143 TACTTGTGTTGCTAGCTTCTCTCACTTATGTTATGTAATAGCGTTTATCCAAAT 202
DB 112730 ACTGATCTTAGTGTGTTGTTGTTTATCTGTGATCTTTCTAAACTCAGTTATTAAGTT 112671
QY 203 TTATGATTTTGTAGGAGAGCAAAATTCGCTGCATGATTAACAGGATCGTCTAA 262
DB 112670 TTAAGAGATTTTATGACATATGTCATACATTTTCTACATGAAAAATCATGTCATCATCA 112611
QY 263 AAA 265
DB 112610 AAA 112608

RESULT 14
US-09-596-002-35
Sequence 35, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Patterson, Chandra
APPLICANT: Berg, Kim, L.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
FILE REFERENCE: PM-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
CURRENT FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 35
LENGTH: 96109
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte template ID No. 6632636 35
PUBLICATION INFORMATION:
US-09-596-002-35

Query Match 7.5%; Score 33.6; DB 3; Length 96109;
Best Local Similarity 51.3%; Pred. No. 9.9; Mismatches 78; Conservative 0; Indels 0; Gaps 0;

QY 31 ATTTAGGTGACTACGATTTAATTTTGGGAAAAACAGGATCGTGGCGGACAGA 90
DB 21998 AGTTTGGTAACGGTCGACCTGACTGATTAATATATATATATCAAGCTTAGAAT 22057
QY 91 CTGTGTGACCGGATATTTAGAGATTTGTGTTAATCAGCGGCTTTATTTACTTGT 150
DB 22058 CATGATTAACGTAAGTTTGGGCTTATTTTCATTAATTAATTCATTTTATGTCAT 22117
QY 151 TTGCTAGTAGCTGCTCACTTATGTTATCGT 182

DB 22118 TTGCCAAGTTTTTACAGCTCACTGTGTGT 22149

RESULT 15
US-09-248-796A-434/C
Sequence 434, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 434
LENGTH: 501
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-434

Query Match 7.4%; Score 33.2; DB 3; Length 501;
Best Local Similarity 56.4%; Pred. No. 1.2; Mismatches 48; Indels 0; Gaps 0;
Matches 62; Conservative 0;

QY 30 AATTTAGGTGACTACTCACTTAAATTTTGGGAAAAACAGGATCGTGGCGGACAG 89
DB 383 AATTTAATACACCTTTGAATTTAATTTTACACAAAAACGTTTAGTGAAC 324
QY 90 ACTTGTGACCGGATTTTGAAGCTTGTGTTAATCAGCGGCTTTTA 139
DB 323 GTTGTGTGCTGAATCTTAACACTCGTGTGTAATCGACGGTGTGA 274

Search completed: February 27, 2006, 06:46:14
Job time : 110.386 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 6563.47 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-3

Perfect score: 1140
Sequence: 1 atgaaaaaagaactgagctt.....gtgacaaactaaatcctaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmb1:*
1: gb_da:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1136.8	99.7	2989	1	AB016245
2	1136.8	99.7	4315	1	AB039950
3	1133.6	99.4	3738	1	AB046355
4	1132	99.3	1143	6	ES0427
5	1132	99.3	3045	6	ES0424
6	1127.2	98.9	8189	1	BS292954
7	1127.2	98.5	200690	1	BSUB0019
8	1122.4	98.5	2775	1	DQ086153
9	540.2	47.4	110000	1	AE017333
10	540.2	47.4	110000	1	CP000002_36
11	332.4	29.2	6896	6	AX433330
12	250.4	22.0	94829	1	AE011191
13	250.4	22.0	94830	1	AE017335
14	250.4	22.0	96231	1	AF188935
15	250	21.9	3244	1	BACCAPABC
16	225.6	19.8	488	6	AX433362
17	103.8	9.1	110000	1	CP000092_3
18	82.6	7.2	110000	1	AE017180_00

C 19	80.8	7.1	110000	1	BA000043_20	Continuation (21 o
C 20	79.4	7.0	110000	1	CP000089_12	Continuation (13 o
C 21	77.8	6.8	249050	1	AL596165	AL596165 Listeria
C 22	77.8	6.8	349980	6	AX417038	AX417038 Sequence
C 23	77.8	6.8	349980	6	AX417041	AX417041 Sequence
C 24	74	6.5	885	6	AX414535	AX414535 Sequence
C 25	74	6.5	2282	6	AX416380	AX416380 Sequence
C 26	73.2	6.4	347456	1	AE017003	AE017003 Bacillus
C 27	73	6.4	110000	1	AP006840_28	Continuation (29 o
C 28	72.6	6.4	280050	1	AL591975	AL591975 Listeria
C 29	72.6	6.4	349980	6	AX641665	AX641665 Sequence
C 30	71	6.2	291073	1	AE017323	AE017323 Listeria
C 31	68.8	6.0	110000	1	AP008934_03	Continuation (4 of
C 32	66.4	5.8	6006	1	BCE17788	AE017270 Bacillus
C 33	66.4	5.8	288437	1	AE017270	Continuation (6 of
C 34	65.8	5.8	110000	1	AP006840_05	BX321864 Nitrosomo
C 35	65.8	5.8	326050	1	BX321864	Continuation (7 of
C 36	65.2	5.7	110000	1	BA000028_06	Continuation (18 o
C 37	63.8	5.6	110000	1	AE017355_17	Continuation (18 o
C 38	62.2	5.5	110000	1	CP000001_17	Continuation (18 o
C 39	60.4	5.3	110000	1	AE017225_17	Continuation (18 o
C 40	60.4	5.3	110000	1	AE017334_17	Continuation (18 o
C 41	60.4	5.3	290525	1	AE017029	AE017029 Bacillus
C 42	56.6	5.0	110000	1	AP006840_09	Continuation (10 o
C 43	56.2	4.9	110000	1	CP000023_01	Continuation (2 of
C 44	56.2	4.9	110000	1	CP000024_01	Continuation (2 of
C 45	56	4.9	930	1	AB125962	AB125962 Bacillus

ALIGNMENTS

RESULT 1
LOCUS AB016245 2989 bp DNA linear BCT 14-OCT-1999
DEFINITION Bacillus subtilis pgsa, pgsb and pgsb genes, complete cds.
ACCESSION AB016245
VERSION AB016245.1 GI:6045071
KEYWORDS pgsa, pgsb, pgsb.
SOURCE Bacillus subtilis
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
AUTHORS Ashiuchi,M., Soda,K. and Misono,H.
TITLE A poly-gamma-glutamate synthetic system of Bacillus subtilis IFO 3336: gene cloning and biochemical analysis of poly-gamma-glutamate produced by Escherichia coli clone cells
JOURNAL Biochem. Biophys. Res. Commun. 263 (1), 6-12 (1999)
PUBMED 10486244

REFERENCE
AUTHORS Ashiuchi,M.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1998) Makoto Ashiuchi, Kochi University, Research Institute of Molecular Genetics; Otsu 200, Monobe, Nankoku, Kochi 783-8502, Japan (E-mail:ashiuchi@img.kochi-u.ac.jp, Tel:0888-64-5215, Fax:0888-64-5109)
FEATURES
source location/Qualifiers
1. 2989
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="IFO 3336"
/sub_species="Bacillus natto"
/db_xref="taxon:1423"
1. 1182
/gene="pgsb"
1. 1182
/gene="pgsb"
/codon_start=1
/transl_table=1
/product="Pgsb"
/protein_id="BA85263.1"
/db_xref="GI:6045072"
/translation="WMLLIACAVILVIGLEKRRHKQNDALPVRVINGIRGKSTV

```

TRLTGILLIAGVKTGKTTGTADAMTYMDPEEKPIKRPQSGISEQKEMRETYE
RGAAIVSECAVNDYQIIFQBELLOANIGVINVLIEDHWGPTLDEIAEAFAT
IPYNGHIVITDSSEYTEEPKOKAKERNTKVIITADNSKITDYLKPEYVLPEDNA
GVALGIDETIARKMLNAPDPDGAAILULISPSEPHGVNPGANDASTNTIK
RVKIGYPTDPIIIMNCRADRVRTQDPADVLPYIEASLILIGETTERIVAYEE
GKIPADKIDHLEIKYSTDEIMELLKSMNHRVYIGVINHGAEBLIEKIHBYKXOLV
S"
gene
1197.1646
/ene="pgsc"
1197.1646
/ene="pgsc"
/codon_start=1
/transl_table=1
/product="Pgsc"
/protein_id="BAAB5264.1"
/db_xref="GI:6045073"
/translation="MFSDLYIALILGVLSLIPAEKTIIVPAGLVPGYIGLVNOR
VFILVLVLSLITVIVIKVYGLSKEMILGRRKFAMLITGVILKIAEDFLPIVPEI
AEFRIGIIVPLIANTIOKQGLITFGSTLLSGATPAIMFYLLI"
1665.2807
/ene="pgsa"
1665.2807
/ene="pgsa"
/codon_start=1
/transl_table=1
/product="Pgsa"
/protein_id="BAAB5265.1"
/db_xref="GI:6045074"
/translation="MKKELSFHEKLLKLTQOKKKTNKHVIAIPVIFVLFAPMAG
KAETPKVYTSDDVLSAFVDIMNGRYVEVTEQKADSIPOYVEPIFRASDVAGN
FENPTYOKNKQADKEIHLQTNKESVYVLKDNMTYVANSNNHMDVQGMQDTLG
EPANONLDIVAGYSISDAKKISYOKNGYIATIGTDTVDGPAKKNTPGVLP
DPEIFPIHSAKKHADIVVQSHHGQDYNDPRDROQLAASDADADIVGHHPH
VLEPIEVNGTIVFISLGNFVFDQGMTRKVSALVOHLKNGGREVEPIDHEAT
PAPVKOSLKOKTIIRELTKDSNFAKVEDKLTEDIDHSDKLSK"

```

ORIGIN

Query Match 99.7%; Score 1136.8; DB 1; Length 2989;

Best Local Similarity 99.8%; Pred. No. 2.9e-276; Mismatches 2; Indels 0; Gaps 0;

```

Matches 1138; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTGTAAGCTGACAAAAACAGCAAAAAAG 60
Db 1665 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTGTAAGCTGACAAAAACAGCAAAAAAG 1724
QY 61 AAAACCAATTAAGCAGTATTTATGCGCATTCGATCGTTTTGTCTTATGTTCCCTTTC 120
Db 1725 AAAACCAATTAAGCAGTATTTATGCGCATTCGATCGTTTTGTCTTATGTTCCCTTTC 1784
QY 121 ATGTGGGGGGGAAAGCGGAAAGCGCGAAGGTCAAAAGCTATTCGAGAGCTACTCTCA 180
Db 1785 ATGTGGGGGGGAAAGCGGAAAGCGCGAAGGTCAAAAGCTATTCGAGAGCTACTCTCA 1844
QY 181 GCCTCATTTGTAGCGCATATTATGATGGGACGCTATGTTGAAAAAGTAACGAGCAAAAA 240
Db 1845 GCCTCATTTGTAGCGCATATTATGATGGGACGCTATGTTGAAAAAGTAACGAGCAAAAA 1904
QY 241 GGGGCGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATGACGA 300
Db 1905 GGGGCGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATGACGA 1964
QY 301 GGAACCTTTGAAAAACCGGTAACTCATCAAAAGATTTTAAACAAGCAGATTAAGAAT 360
Db 1965 GGAACCTTTGAAAAACCGGTAACTCATCAAAAGATTTTAAACAAGCAGATTAAGAAT 2024
QY 361 CATCTGACAGCAATTAAGAAATCAGTAAAGTCTTGAAGATATGAATTTTCAGGTTCTC 420
Db 2025 CATCTGACAGCAATTAAGAAATCAGTAAAGTCTTGAAGATATGAATTTTCAGGTTCTC 2084
QY 421 AACAGCGCCAAACACGCAATGATTAACGCGCTTCAGGCGATGAAGATACGCTTGA 480
Db 2085 AACAGCGCCAAACACGCAATGATTAACGCGCTTCAGGCGATGAAGATACGCTTGA 2144

```

```

QY 481 GAATTTCCGAAAGCAAAACCTTGATATGTTGAGCGGGATACAGCTTAAGTATGCGAAA 540
Db 2145 GAATTTCCGAAAGCAAAACCTTGATATGTTGAGCGGGATACAGCTTAAGTATGATGCGAAA 2204
QY 541 AAGAAATTTTCGTAACGAAAGTCAACGGGGTAAACGATTGCAACGCTTGCTTTACCGAT 600
Db 2205 AAGAAATTTTCGTAACGAAAGTCAACGGGGTAAACGATTGCAACGCTTGCTTTACCGAT 2264
QY 601 GTGTCCGGGAAAGGTTTCGGGGCTTAAAGAAATACGCGGGGCGTCCGCGAGATCCT 660
Db 2265 GTGTCCGGGAAAGGTTTCGGGGCTTAAAGAAATACGCGGGGCGTCCGCGAGATCCT 2324
QY 661 GAAATCTTATCCCTTATGATTTTCAAGAGCAAAAAACATGCTGACATTTGTTGTGACG 720
Db 2325 GAAATCTTATCCCTTATGATTTTCAAGAGCAAAAAACATGCTGACATTTGTTGTGACG 2384
QY 721 TCACACTGGGGCCCAAGGTATGACATATGATCCAAACGACCGCCGACGCTTGCAAGA 780
Db 2385 TCACACTGGGGCCCAAGGTATGACATATGATCCAAACGACCGCCGACGCTTGCAAGA 2444
QY 781 GCCATGCTGATGCGGGAGGCTGACATCATGTCGGGCATCATCCGACGCTTTAGAACG 840
Db 2445 GCCATGCTGATGCGGGAGGCTGACATCATGTCGGGCATCATCCGACGCTTTAGAACG 2504
QY 841 ATTAAGTATATTAACGAAACCGTCAATTTTCTACAGCTTCGCAACTTTGTCTTGAACAA 900
Db 2505 ATTAAGTATATTAACGAAACCGTCAATTTTCTACAGCTTCGCAACTTTGTCTTGAACAA 2564
QY 901 GCGTGAACGAGAACAGACAGACAGTGCATCTGTTGAGTATCACTGAAAGAAATGGAACA 960
Db 2565 GCGTGAACGAGAACAGACAGACAGTGCATCTGTTGAGTATCACTGAAAGAAATGGAACA 2624
QY 961 GCGCGCTTTGAAGAGACACGATGATATCATGATCAATGACGACACCTGACCTGTGAAAAA 1020
Db 2625 GCGCGCTTTGAAGAGACACGATGATATCATGATCAATGACGACACCTGACCTGTGAAAAA 2684
QY 1021 GACAGCCTTAAACGAAACCATTAATTCGCGAATCTGACGAAAGACTTAATTTGCTTGG 1080
Db 2685 GACAGCCTTAAACGAAACCATTAATTCGCGAATCTGACGAAAGACTTAATTTGCTTGG 2744
QY 1081 AAGTAGAAGACGAAACCTGACGTTTGATATGATCATATGATCAAACTTAATCTTAA 1140
Db 2745 AAGTAGAAGACGAAACCTGACGTTTGATATGATCATATGATCAAACTTAATCTTAA 2804

```

RESULT 2

AB039950 LOCUS 4315 bp DNA linear BCT 12-SEP-2000

DEFINITION Bacillus subtilis capB, capC, capA, ywC genes, complete cds.

ACCESSION AB039950

VERSION AB039950.1 GI:10119860

KEYWORDS CapA; CapC; CapB.

SOURCE

ORGANISM Bacillus subtilis

REFERENCE 1 (bases 1 to 4315)

AUTHORS Tran, L.P. and Itoh, Y.

TITLE Nucleotide sequence of the capBCA operon in Bacillus subtilis (natto)

JOURNAL Published Only in Database (2000)

REFERENCE 2 (bases 1 to 4315)

AUTHORS Tran, L.P. and Itoh, Y.

TITLE Direct Submission

JOURNAL Submitted (08-MAR-2000) Yoshifumi Itoh, National Food Research Institute, Applied Microbiology; Kannonkai 2-1-2, Tsukuba, Ibaraki 305-8642, Japan (E-mail: yosifumi@nfi.affrc.go.jp, Tel.: +81-298-38-8075, Fax: +81-298-38-7996)

FEATURES

source

1..4315 location/qualifiers

/organism="Bacillus subtilis"

/mol_type="genomic DNA"

/strain="Aeahikawa"

/db_xref="taxon:1423"

```

gene
941..2122
/gene="capB"
CDS
941..2122
/gene="capB"
/codon_start=1
/transl_table=11
/product="CapB"
/protein_id="BAB13484.1"
/db_xref="GI:10119861"
/translation="MWLLIACAVILVIGILEKRRHQNIDALPVRVNIINGIRKSTV
RLTGLIAGYKTVKGTGDAMIYMDPEEKPIKRGQPMIGOEKVMRETV
RGANAIVSECAVNDYQIIOBELLOANIYIVNFDHDMVMPITDEIAEATAT
IPNCHIVITDSSEVPEPKOKAKERNTVITIANDKITDEYIKREPVWFPNLSAL
GVAOLGIDEETAFKMLNAPDPGAMKILPILISBEQHPFNGAANDASTINMK
RVKEIGPTDPIIIMNCRADVKRTOQFADVLPYIASELILIGETHEPIVAYEE
GKIPADKHLDELKSTDEIMELKRMHNRVIYGVNTHGAEPILIEKHXYQLV
S"
2137..2586
/gene="capC"
CDS
2137..2586
/gene="capC"
/codon_start=1
/transl_table=11
/product="CapC"
/protein_id="BAB13485.1"
/db_xref="GI:10119862"
/translation="MFGSDLYIALILGVLLSIFAEKGTIVPAGLVPGYGLVENQP
VETILVLSLTLYIVKYGLSKFMILVGRKFAAMLITGLIKIAFPFLYPIVPEI
AEFRIGIIVPLINANTIQKGLITTFSGTLLSGTFAIMFYVYLI"
2605..3747
/gene="capA"
CDS
2605..3747
/gene="capA"
/codon_start=1
/transl_table=11
/product="CapA"
/protein_id="BAB13486.1"
/db_xref="GI:10119863"
/translation="MKKELSFHEKLKLTQKQKKTINKHPIAIPVFLMEAFMWAG
KAETPKVATSDVLSASPVGDIMMGRVYEKTEQKADSIPOYVEPIFRAADYVAG
PENPYOKNKOAKKEIHLQTNKESVYKLKDMNFTVANSANHAMDVCGMKDITG
EBAKQNDIVGAGYSISDAKKISYKQNVITATLGFDPVSGKGAACKQNGYLP
DEIPIPMISEAKHADIVVQSHWQEDNDPNDRQDLAAMSADAGDITVGHHPH
VLEPIEVNGTVIFYSLGNFVPOGWTTRDSALVOYHLKKGKGTGFETPIDIHEAT
PAPVKDLSLKQTIIRLTKDSNFAWKVEDKLTFDIDHSDLKSKR"
3762..3929
/gene="ywcC"
CDS
3762..3929
/gene="ywcC"
/codon_start=1
/transl_table=11
/protein_id="BAB13487.1"
/db_xref="GI:10119864"
/translation="MKFVKAIMPFAVALVPMFMSAFKENDQLTDEKQKIDEMNKI
QOQEEPVNANK"
ORIGIN
Query Match 99.7%; Score 1136.8; DB 1; Length 4315;
Best Local Similarity 99.8%; Pred. No. 2.9e-278;
Matches 1138; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTCTAAAGCTGACAAAACGACAAAAAAG 60
Db 2605 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTCTAAAGCTGACAAAACGACAAAAAAG 2664
Qy 61 AAAACCAATAGACGATTTATTTGCGATTCGATCGTTTTTGTCTTATGTTGCTTTC 120
Db 2665 AAAACCAATAGACGATTTATTTGCGATTCGATCGTTTTTGTCTTATGTTGCTTTC 2724
Qy 121 ATGTGGCGGGGAAAGCGGAAAGCGGAAAGCTCAAAAGCTATTCGACGAGTACTCTCA 180
Db 2725 ATGTGGCGGGGAAAGCGGAAAGCGGAAAGCTCAAAAGCTATTCGACGAGTACTCTCA 2784

```

```

Qy 181 GCCTCATTTGAGCGCATATTATGATGGAGCGTATGTTGAAAAAGTAACGACCAAAA 240
Db 2785 GCCTCATTTGAGCGCATATTATGATGGAGCGTATGTTGAAAAAGTAACGACCAAAA 2844
Qy 241 GGGGACAGCAGTATTTTCAATATGTTGAACCGATCTTTAAGCTCGGATTTATGTACCA 300
Db 2845 GGGGACAGCAGTATTTTCAATATGTTGAACCGATCTTTAAGCTCGGATTTATGTACCA 2904
Qy 301 GGAACCTTTGAAACCCGGTAACTATCAAAAAGATTATAACAAGCAGATAAGAGATT 360
Db 2905 GGAACCTTTGAAACCCGGTAACTATCAAAAAGATTATAACAAGCAGATAAGAGATT 2964
Qy 361 CATCTGACAGCAATTAAGAAATCAAGTAAGTCTTGAAGATATGATTAACGGTCTC 420
Db 2965 CATCTGACAGCAATTAAGAAATCAAGTAAGTCTTGAAGATATGATTAACGGTCTC 3024
Qy 421 AACAGCGCAACCAACGCAATGATTAACGGCTTCAAGGCAATGAAGATACGTTTGA 480
Db 3025 AACAGCGCAACCAACGCAATGATTAACGGCTTCAAGGCAATGAAGATACGTTTGA 3084
Qy 481 GAATTTGCAAGCAAAACCTTGATATCGTTGAGAGGGGATACAGCTTAAGTATGATGAAA 540
Db 3085 GAATTTGCAAGCAAAACCTTGATATCGTTGAGAGGGGATACAGCTTAAGTATGATGAAA 3144
Qy 541 AAGAAATTTTGTATGACAGAAAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 600
Db 3145 AAGAAATTTTGTATGACAGAAAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 3204
Qy 601 GTGTCCGGGAAAGGTTTGGCGGCTAAAGAAATACGCGGGGCTGCTCCCGCAGATCT 660
Db 3205 GTGTCCGGGAAAGGTTTGGCGGCTAAAGAAATACGCGGGGCTGCTCCCGCAGATCT 3264
Qy 661 GAATCTTCATCTCCATATGATTTCAAGACGCAAAAAACATGCTGACATGTTGTTGACG 720
Db 3265 GAATCTTCATCTCCATATGATTTCAAGACGCAAAAAACATGCTGACATGTTGTTGACG 3324
Qy 721 TCACACTGGGGCCAGAGATGACAAATGATCCAAACGACCGGCAAGCTTGCAAGA 780
Db 3325 TCACACTGGGGCCAGAGATGACAAATGATCCAAACGACCGGCAAGCTTGCAAGA 3384
Qy 781 GCCATGCTGATGCGGAGCTGACATCATGCTGCGCATCATCGCACTCTTAAACCG 840
Db 3385 GCCATGCTGATGCGGAGCTGACATCATGCTGCGCATCATCGCACTCTTAAACCG 3444
Qy 841 ATTGAAGTATATACGGAACCGTCAATTTCTACAGCTCGGCAACTTTGCTTTAGACCA 900
Db 3445 ATTGAAGTATATACGGAACCGTCAATTTCTACAGCTCGGCAACTTTGCTTTAGACCA 3504
Qy 901 GGCTGACGAGAAACAAGAGACAGTGCATGTTCAAGTATCACCTGAAGAAATGAGACA 960
Db 3505 GGCTGACGAGAAACAAGAGACAGTGCATGTTCAAGTATCACCTGAAGAAATGAGACA 3564
Qy 961 GCGCGCTTTGAAGTACACCGATCGATATCATGAAAGCACACTGACCTGTGAAAAA 1020
Db 3565 GCGCGCTTTGAAGTACACCGATCGATATCATGAAAGCACACTGACCTGTGAAAAA 3624
Qy 1021 GACAGCTTTAAACAGAAACCATATTTGCGGAACTGACAAAGACTTAATTTGCTTGG 1080
Db 3625 GACAGCTTTAAACAGAAACCATATTTGCGGAACTGACAAAGACTTAATTTGCTTGG 3684
Qy 1081 AAGTGAAGACGGAATAATGACGTTGATATGATCATAGTGAACAACTAAATCTAAA 1140
Db 3685 AAGTGAAGACGGAATAATGACGTTGATATGATCATAGTGAACAACTAAATCTAAA 3744

```

RESULT 3
 AB046355 3738 bp DNA linear BCT 11-JAN-2002
 LOCUS Bacillus subtilis ywcC, ywcA, ywcB, ywcC genes, complete cds.
 DEFINITION
 ACCESSION AB046355
 VERSION AB046355.1 GI:13591556
 KEYWORDS
 SOURCE Bacillus subtilis

ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE 1 Urushibata, Y., Tokuyama, S. and Tahara, Y.
TITLE Characterization of the Bacillus subtilis ywsc gene, involved in
gamma-polyglutamic acid production
JOURNAL J. Bacteriol. 184 (2), 337-343 (2002)
PUBMED 11751809

AUTHORS 2 (bases 1 to 3738)
Tahara, Y. and Urushibata, Y.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2000) Yaeutaka Tahara, Shizuoka University,
Faculty of Agriculture, 836 Ohya, Shizuoka city, Shizuoka 422-8529,
Japan (E-mail: acyutah@agr.shizuoka.ac.jp,
Tel: 81-54-238-4878 (ex. 7808), Fax: 81-54-237-3028)

FEATURES Location/Qualifiers
1. 3738
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/db_xref="taxon:1423"
497..1678
/gene="ywsc"
497..1678
/gene="ywsc"
/codon_start=1
/transl_table=1
/protein_id="BAB40948.1"
/db_xref="GI:13591557"
/translation="WMLLIACAVILVIGLEKRRHQKIDALPVRVNINGIRKSTV
TRLTTGILIEAGYKTVGKTGTDAIMIYWDPEEKPIRKQGPVIGQKMEVTE
RGNAAVISECAVNDYOIIFQEBELQNIQIVNVLEHNDVNGPFLDEAFATAT
IPVNGHLYITDSEYTEFQKAKERTNTVLIADNSKRTDEYLRKEVWVPDNLAL
GVAALGIDEETAFKGMNAPDGGAMILPLISSEBGFHNGPAAVADASTINIK
RVKEGYTDDPIITMNCRAPRVDTQCPADVPIYASLLIIGEOEPIYAYEE
GKIPADKIDLEYSIDELIMELKKRMKNRVLYGVNHGAELPLEIKTHYKAKOLV
S"
1693..2142
/gene="ywca"
1693..2142
/gene="ywca"
/codon_start=1
/transl_table=1
/protein_id="BAB40949.1"
/db_xref="GI:13591558"
/translation="MFGSDLYALILIGVLSLIPAEKTIYAGVIVPVGILVFNOP
VFILVLVSLIIVYKRGSKEMILYGRKKPAMLTGIVLKAAPFLYPIYPREI
AEFRGIIIVPEGLANTIQKQGLTTFESTLLSGRAFMFVYLLI"
2161..3303
/gene="ywcb"
2161..3303
/gene="ywcb"
/codon_start=1
/transl_table=1
/protein_id="BAB40950.1"
/db_xref="GI:13591559"
/translation="MKELSPHEKLLKTLKQKKKTNKGVFIAPVIFVLEAFMMAG
KAEPKVTYDDVDASFPVDIMMGVVEKTEBQKDSIROYEPIFRASDVAGN
FENPTVYKAKQADKEIHLOTNKESYVULDNMTFVANSNNHMDVQVGMQDTG
EPATKNDIVGAGTSLDAKKKISTQKNGVITATLGTTDVSQKFAKKAPGMLPA
DPELIPMISEAKKHADIIVVQSHWGQDYDNDPNRQQLARMAASDAADAIIVGHHP
VLEPIEVNGTVIFYSIGNFVFDQGGTTRDSALVOYHKKKGTGRFEVTPDIHEAT
PAPVKDLSLKQKTIIRLETKDSNFAWKVEDGLTFDIDHSKLSK"
3318..3485
/gene="ywcc"
3318..3485
/gene="ywcc"
/codon_start=1
/transl_table=1
/protein_id="BAB40951.1"
/db_xref="GI:13591560"
/translation="WKFVKAIMPVAVAVIVFMNSAFKENDQLTDQEKQKIDMEKNKI
OOOEPPVANK"

gene
CDS

gene
CDS

gene
CDS

ORIGIN

Query Match 99.4%; Score 1133.6; DB 1; Length 3738;
Best Local Similarity 99.6%; Pred. No. 1,9e-277;
Matches 1136; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAAAGAACTGACCTTTCATGAAAGCTGTAAAGCTGACAAACGCAAAAAAG 60
2161 ATGAAAAAGAACTGACCTTTCATGAAAGCTGTAAAGCTGACAAACGCAAAAAAG 2220
61 AAACCAATTAAGCAAGTATTTATGCGATTCGATTCGATTCGATTCGATTCGATTC 120
2221 AAACCAATTAAGCAAGTATTTATGCGATTCGATTCGATTCGATTCGATTCGATTC 2280
121 ATGCGCGCGGAAAAAGCGAAACGCCAAGCTCAAAAGTATTCGACGACGATCTCA 180
2281 ATGCGCGCGGAAAAAGCGAAACGCCAAGTCAAAAGTATTCGACGACGATCTCA 2340
181 GCTCATTTTGTAGCGCATTTATGATGAGACGCTATGTTGAAAAAGTAAACGACAAA 240
2341 GCTCATTTTGTAGCGCATTTATGATGAGACGCTATGTTGAAAAAGTAAACGACAAA 2400
241 GGGGACAGCATTTTTCATATGATGAGACGCTATGAGCGCTGAGATTATGATGAC 300
2401 GGGGACAGCATTTTTCATATGATGAGACGCTATGAGCGCTGAGATTATGATGAC 2460
301 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTTATTAACAGCAGATTAAGATTT 360
2461 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTTATTAACAGCAGATTAAGATTT 2520
361 CATTCGACAGCAATTAAGCAATCATGTAAGTCTTGAAGATTAATTTACGCTTCTC 420
2521 CATTCGACAGCAATTAAGCAATCATGTAAGTCTTGAAGATTAATTTACGCTTCTC 2580
421 AACGCCCAACCAACGCAATGATTAAGCGGCTTGAAGGCAATTAAGATTAAGCTTGA 480
2581 AACGCCCAACCAACGCAATGATTAAGCGGCTTGAAGGCAATTAAGATTAAGCTTGA 2640
481 GAATTCGAAAGCAAAACCTTATGATTCGTTGAGCGGATTAAGCTTAAGTATGCGAAA 540
2641 GAATTCGAAAGCAAAACCTTATGATTCGTTGAGCGGATTAAGCTTAAGTATGCGAAA 2700
541 AAGAAATTTTCGACGAAAGTCAACGCGGATTAAGTCAACGCTTGCCTTACCGAT 600
2701 AAGAAATTTTCGACGAAAGTCAACGCGGATTAAGTCAACGCTTGCCTTACCGAT 2760
601 GTGTCGGGAAAGTTTCGCGGCTTAAAGAAATTCGCGCGGCGTCCGCGGACGATCCT 660
2761 GTGTCGGGAAAGTTTCGCGGCTTAAAGAAATTCGCGCGGCGTCCGCGGACGATCCT 2820
661 GAATCTTCATCCTTATGATTCGAAAGCAAAACATGCTGACATTTGTTGTGTCAG 720
2821 GAATCTTCATCCTTATGATTCGAAAGCAAAACATGCTGACATTTGTTGTGTCAG 2880
721 TCACATCGGGGCAAGATGATGACATGATTCAAACGACCGCCAGCGGACCTTGCAAGA 780
2881 TCACATCGGGGCAAGATGATGACATGATTCAAACGACCGCCAGCGGACCTTGCAAGA 2940
781 GCCATGCTGATGCGGAGCGATCATGCTGCGGCATCATCTGCGACGCTTGAACCG 840
2941 GCCATGCTGATGCGGAGCGATCATGCTGCGGCATCATCTGCGACGCTTGAACCG 3000
841 ATTGAAGATTAACGAAACCGCATTTTTCACAGCTCGGCACTTGTCTTGAACAA 900
3001 ATTGAAGATTAACGAAACCGCATTTTTCACAGCTCGGCACTTGTCTTGAACAA 3060
901 GCGTGAACGAAACGAGACGATGCACTGCTCAGTATCACTGAAGAAATGGAACA 960
3061 GCGTGAACGAAACGAGACGATGCACTGCTCAGTATCACTGAAGAAATGGAACA 3120
961 GCGCGCTTTGAAGGACACCGATGATATCACTGAAGCACTGCACTGTGTAAGAAA 1020
3121 GCGCGCTTTGAAGGACACCGATGATATCACTGAAGCACTGCACTGTGTAAGAAA 3180

QY 1021 GACAGCCTTTAAACAGAAAACCTATTATGCGCACTGAGAAAGACTTAATTTGCTGG 1080
 DB 3181 GACAGCCTTTAAACAGAAAACCTATTATGCGCACTGAGAAAGACTTAATTTGCTGG 3240
 QY 1081 AAAGTAGAAGACGAAAACCTGAGCTTTGATTTGATTCATAGTACAACTAAATCTTAA 1140
 DB 3241 AAAGTAGAAGACGAAAACCTGAGCTTTGATTTGATTCATAGTACAACTAAATCTTAA 3300
 RESULT 4
 E50427 1143 bp DNA linear PAT 31-JAN-2002
 LOCUS E50427
 DEFINITION Process for producing poly-gamma-glutamic acid.
 ACCESSION E50427
 VERSION E50427.1 GI:18629415
 KEYWORDS JP 2001017182-A/4.
 SOURCE Bacillus subtilis
 ORGANISM Bacillus subtilis
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (bases 1 to 1143)
 AUTHORS Ashiuchi, M., Misono, H. and Soda, K.
 TITLE Process for producing poly-gamma-glutamic acid
 JOURNAL Patent: JP 2001017182-A 4 23-JAN-2001;
 NAGASE ECO LTD
 OS Bacillus subtilis (hay bacillus) IFO 3336
 PN JP 2001017182-A/4
 PD 23-JAN-2001
 PR 09-JUL-1999 JP 1999196335
 PC MAKOTO ASHIUCHI, HARUO MISONO, KENJI SODA
 C12N15/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
 C12N9/00, C12N9/90,
 CC C12P13/14, C12N15/00, C12N5/00
 FH Key Location/Qualifiers
 FT source 1. 1143
 FT /organism="Bacillus subtilis"
 FT IFO 3336
 FT Location/Qualifiers
 1. 1143
 /organism="Bacillus subtilis"
 /mol_type="genomic DNA"
 /db_xref="taxon:1423"
 ORIGIN
 Query Match 99.3%; Score 1132; DB 6; Length 1143;
 Best Local Similarity 99.6%; Pred. No. 4.7e-277;
 Matches 115; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 ATGAAAAAAGAACTGAGCTTTGATGAAAAGCTGCTAAAGCTGACAAAACGCAAAAAAG 60
 DB 1 ATGAAAAAAGAACTGAGCTTTGATGAAAAGCTGCTAAAGCTGACAAAACGCAAAAAAG 60
 QY 61 AAAACCAATPAGACGATTTATTTGCGATTCGATGTTTTTGTCTTATGTTGCTTTC 120
 DB 61 AAAACCAATPAGACGATTTATTTGCGATTCGATGTTTTTGTCTTATGTTGCTTTC 120
 QY 121 ATGTGGCGGGAAGAAAGCGGAAAGCGGAAAGTCAAAACGTAATCTGACGATCTCA 180
 DB 121 ATGTGGCGGGAAGAAAGCGGAAAGCGGAAAGTCAAAACGTAATCTGACGATCTCA 180
 QY 181 GCCTCATTTGTAGCGCATATTATGATGGAAGCGTATGTTGAAAAAGTAAACGAGCAAAA 240
 DB 181 GCCTCATTTGTAGCGCATATTATGATGGAAGCGTATGTTGAAAAAGTAAACGAGCAAAA 240
 QY 241 GGGGCGAGACGATTTTCAATATGTTGAACCGATCTTTTGAAGCTTGGATTTATGTA 300
 DB 241 GGGGCGAGACGATTTTCAATATGTTGAACCGATCTTTTGAAGCTTGGATTTATGTA 300
 QY 301 GGAACCTTTGAAAAACCGGTACCTATCAAAAAGAAATATATAACAAAGAGATTAAGAGATT 360
 DB 301 GGAACCTTTGAAAAACCGGTACCTATCAAAAAGAAATATATAACAAAGAGATTAAGAGATT 360

QY 361 CATCTGACAGCAATTAAGAAATCACTGAAAGCTTGAAGATATGAAATTTACGGTTCTC 420
 DB 361 CATCTGACAGCAATTAAGAAATCACTGAAAGCTTGAAGATATGAAATTTACGGTTCTC 420
 QY 421 AACAGCGCAACCAACACGCAATGATTAAGCGCTTCAAGGCAATGAAGATATCGTTTGA 480
 DB 421 AACAGCGCAACCAACACGCAATGATTAAGCGCTTCAAGGCAATGAAGATATCGTTTGA 480
 QY 481 GAATTTGGAAAGCAAAACCTTGATATCGTTGAGCGGATACAGCTTAAGTATGATCGAAA 540
 DB 481 GAATTTGGAAAGCAAAACCTTGATATCGTTGAGCGGATACAGCTTAAGTATGATCGAAA 540
 QY 541 AAGAAATTTTGTATCCAGAAAGTCAACCGGGTAAACGATTGCAACGCTTTACCGAT 600
 DB 541 AAGAAATTTTGTATCCAGAAAGTCAACCGGGTAAACGATTGCAACGCTTTACCGAT 600
 QY 541 AAGAAATTTTGTATCCAGAAAGTCAACCGGGTAAACGATTGCAACGCTTTACCGAT 600
 DB 541 AAGAAATTTTGTATCCAGAAAGTCAACCGGGTAAACGATTGCAACGCTTTACCGAT 600
 QY 601 GTGTCCGGGAAAGGTTTTCGGGCTTAAAGAAATACCGCGGCGTCTCCCGCAGATCTT 660
 DB 601 GTGTCCGGGAAAGGTTTTCGGGCTTAAAGAAATACCGCGGCGTCTCCCGCAGATCTT 660
 QY 661 GAAATCTTCAATCCCTATGATTTTCAAGACGGAAGAAACATGCTGACATTTGTTGGCAG 720
 DB 661 GAAATCTTCAATCCCTATGATTTTCAAGACGGAAGAAACATGCTGACATTTGTTGGCAG 720
 QY 721 TCACACTGGGGCCAGAGATGATGACATGATCCAAACGACCGCCAGCGCTTGCAGAA 780
 DB 721 TCACACTGGGGCCAGAGATGATGACATGATCCAAACGACCGCCAGCGCTTGCAGAA 780
 QY 781 GCCATGCTGATGCGGAGCTGACATCATGCTGCGCATCATGCGCAGCTTTAGAACCG 840
 DB 781 GCCATGCTGATGCGGAGCTGACATCATGCTGCGCATCATGCGCAGCTTTAGAACCG 840
 QY 841 ATTGAAGTATTAACGGAACCGCTATTCTTCAACCGCTGGAACCTTTGCTTTCAGCA 900
 DB 841 ATTGAAGTATTAACGGAACCGCTATTCTTCAACCGCTGGAACCTTTGCTTTCAGCA 900
 QY 901 GGCTGAGACGAAGAACAGAGACAGTGCATGCTGCTTCACTGATCACTGAAAGAAATGGAACA 960
 DB 901 GGCTGAGACGAAGAACAGAGACAGTGCATGCTGCTTCACTGATCACTGAAAGAAATGGAACA 960
 QY 961 GGCGGCTTTGAAGTGAACCGATGATATCATGGAAGGACACCTGACCTGTGAAAAA 1020
 DB 961 GGCGGCTTTGAAGTGAACCGATGATATCATGGAAGGACACCTGACCTGTGAAAAA 1020
 QY 1021 GACAGCCTTTAAACAGAAAACCTATTATGCGCACTGAGAAAGACTTAATTTGCTGG 1080
 DB 1021 GACAGCCTTTAAACAGAAAACCTATTATGCGCACTGAGAAAGACTTAATTTGCTGG 1080
 QY 1081 AAAGTAGAAGACGAAAACCTGAGCTTTGATTTGATTCATAGTACAACTAAATCTTAA 1140
 DB 1081 AAAGTAGAAGACGAAAACCTGAGCTTTGATTTGATTCATAGTACAACTAAATCTTAA 1140
 RESULT 5
 E50424 3045 bp DNA linear PAT 31-JAN-2002
 LOCUS E50424
 DEFINITION Process for producing poly-gamma-glutamic acid.
 ACCESSION E50424
 VERSION E50424.1 GI:18629412
 KEYWORDS JP 2001017182-A/1.
 SOURCE Bacillus subtilis
 ORGANISM Bacillus subtilis
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (bases 1 to 3045)
 AUTHORS Ashiuchi, M., Misono, H. and Soda, K.
 TITLE Process for producing poly-gamma-glutamic acid
 JOURNAL Patent: JP 2001017182-A 1 23-JAN-2001;
 NAGASE ECO LTD
 OS Bacillus subtilis (hay bacillus) IFO 3336
 PN JP 2001017182-A/1
 PD 23-JAN-2001
 PR 09-JUL-1999 JP 1999196335

DNHMGPTLDEIAEFTATIPYNGHVLITSEYTEPEFKOKAKERNTKVIADNSKIT
DEYLRKEFYMFPPNASCIALVAQALGIDEFATARKMANNAPDDGAMRILPLISPEP
GHFVNGFANDASSLTINMKRVKEIGYPTDPTIIMNCRADVRBTOQFANDVPIYE
ASELILIGETTEPIVKAEBGKIPADKLDLEKYSTDEIMELKMSMNRVIYGVGNI
HGAAPLEKIEHYKQVLS"
join(1069, .1075, 1084, .1533)
/gene="ywca"
1069, .1075
/gene="ywca"
1084, .1533
/gene="ywca"
/note="product highly similar to Bacillus anthracis CapC
protein"
/codon_start=1
/transl_table=11
/protein_id="CAB07468.1"
/db_xref="GI:1894766"
/db_xref="GOA:P96737"
/db_xref="UniProt/TREMBL:P96737"
/translation="MFGSDLYALILGVLSLIPAEKGIYVAGIVPGYLGIVNQ
VFILVLVSLTIVIVKGLSKFMILYGRKRPAMLTGLVLKIAFDFLPYVEFI
AEFRGIGIIVPLIANTIOKQGLITTFSTLLSGATPAIMFVYLI"
1538, .2694
/gene="ywcb"
1538, .1544
/gene="ywcb"
1552, .2694
/gene="ywcb"
/note="product highly similar to Bacillus anthracis CapA
protein"
/codon_start=1
/transl_table=11
/protein_id="CAB07469.1"
/db_xref="GI:1894767"
/db_xref="UniProt/TREMBL:P96738"
/translation="MKKLSFHEKLIKLTROOKKXKTNKHVFIPIPIVPLMFAPWAG
KAETPKVTSIDVTSASFVGDIMMGRYVEKTEQKADSIFQYIEPIPRADYVAGN
FENPVYQKNYQADKEIHLQTNKESVYKAKMNFVLANSANNHNDVGVQMKOTLG
BEAKQNLIDVAGYSLSDAKKISYQKVNQVITATLGFVDGSKGFAAKKNTGVLP
DEIFIPMISEAKKADIVVQSHNGQEDNDPNROQLARMSDAGADIIIVGHHP
VLEPIEVNGTIVFISLGNFVFDGWTTRDSALVOYHLKKKGTGRFETPIDHEAT
-PAPVKKDSLKQKTIIRBELTKDSNFAKVEDGKLTIDHSDLKSK"
2697, .2914
/gene="ywcb"
2697, .2702
/gene="ywcb"
2709, .2876
/gene="ywcb"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAB07470.1"
/db_xref="GI:1894768"
/translation="MKFVKAIWPFVAVALVFMMSAFKFNDDLTDEKQKIMENAKI
QOQEBPVNANK"
2880, .2914
/gene="ywcb"
3000, .4258
/gene="ywcd"
3000, .3007
/gene="ywcd"
3017, .4258
/gene="ywcd"
/note="product similar to E. coli NUPC protein and to
Listeria species pto-related proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB07476.1"
/db_xref="GI:1894769"
/db_xref="GOA:P96740"
/db_xref="UniProt/Swiss-Prot:P96740"
/translation="MNTLAWKKFLVAVITCLVPIWTKAETAEADTSSSELTIVSEAK

NILGYQYKGEETPEKEGDPBSGLIQVFSKADIIHLPRS VNDQYKIGTAVPENIKPCD
ILPDKKSGTGYTTHDALYIGDQWHSQTOSKGVITNNKSSYWSGTYIGARILAA
DPAADVPVOAEKAYIGVPPYEGSGTSPSGFDGSLQVYVFOALGYLERSAQOM
AVGKAPAPKINIKPDVYVPSNTYVKTGISHAGIVYAGAPRTIOASSEKVTISYLEDY
KSKWTGIRRDNDTLIPAKENPLVSEATIIYVGEVPYKQGVPTPEPDRAGFOVYQY
AGISLPYATSVQVNAKRIEKADLPKPDIVFPSTSLNPSITYIGNGVAVHTLSNGVT
ITNNSTSTVMKDYKAGSIRVO"
4259, .4293
complement(4292, .5170)
/gene="ywcb"
complement(4292, .5152)
/gene="ywcb"
/note="product similar to Bacillus subtilis yxh and yce
proteins and to E. coli yidA protein"
/codon_start=1
/transl_table=11
/protein_id="CAB07471.1"
/db_xref="GI:1894770"
/db_xref="GOA:P96741"
/db_xref="UniProt/TREMBL:P96741"
/translation="MKCIAIDGLTLNKESVISAENREAIKRAVDAGILVTICGRA
TFDVKALUDDADITITANGSTTHDTGRILSRILMOAGKALADVLSSKNIEEY
TDDHILSPFDEAVLHAEIDLKSNAPNEQTDILMOGANTOPKQGIPIPIHISVP
GGENIYKLCFSPMDKLOKAKELKHKRLAQTSQGHITELIPASGKRALTKLA
DIYGIETQDIYAIGDSPDLSPFEVAGHRIAMENAIDELKESFTVTKSNDENGVAIF
IDOLSGOYA"
complement(5163, .5170)
/gene="ywcb"
5277, .5305
/gene="ywcb"
5277, .5305
/gene="ywcb"
complement(5308, .6291)
/gene="ywcb"
/pseudo
complement(5308, .6276)
/gene="ywcb"
/note="an in frame TGA stop codon is found at position
5599 to 5601
translation similar to Bacillus subtilis lycR protein"
/pseudo
/codon_start=1
/transl_table=11
complement(6285, .6291)
/gene="ywcb"
6594, .8112
/gene="ywcb"
6594, .6601
/gene="ywcb"
6609, .7982
/gene="ywcb"
/note="product highly similar to metabolite transport
proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB07473.1"
/db_xref="GI:1894771"
/db_xref="GOA:P96742"
/db_xref="UniProt/TREMBL:P96742"
/translation="MKKOSNIWLYFFGALGALYGDYGVISGAILFMKKEBGLNAFT
BGLVSSILVGAIIIGSGAAGKLTDRFGKKAISLNMOTVGLISYIVNYFADAEA
RIILGLANGSTSTIYPLVLSIELAPKRGKALSNOMITVGLISYIVNYFADAEA
WRMWGLAIVPSLLILIGLWEPSPRNLFTNGESSKAKKIIIEKRGTDIDOEIHI
KEAKQODBGKLELFDPMWRPLIAGLGIAFLQFTGNTTIIYAPKFTTNGFGNSA
SLIGTIGTGVAVMTLVAIKIIDIKGKPLLFNAGAVSLIVLALNVEFNNTPA
ASMTTIVICGIVIVFAVSWGVVWMLPELPLHVRIGTGVSITLMHVTGLIYSLT
YVIMEAIGISLPLIYALIGIAFLFRFKVTEKRSLEIEIDDLRDKNGGGAAG
KOQTVGT"
7981, .8112
/gene="ywcb"
complement(8019, .8189)
/gene="gerBC"
complement(8019, .>8189)


```

/gene="gezBC"
/codon_start=1
/translate=1
/product="gezBC"
/protein_id="CAB07474.1"
/db_xref="GI:1894772"
/db_xref="GOA:P96743"

Query Match      98.9%   Score 1127.2; DB 1; Length 8189;
Beet Local Similarity 99.3%   Pred. No. 8.4e-276;
Matches 1132; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 1 ATGAAAAAGAACTGAGCTTTCATGAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 60
Db 1552 ATGAAAAAGAACTGAGCTTTCATGAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 1611

Oy 61 AAAACCAATGACGATTTTATGTCCTCCGATCGTTTGTCTTATGTTGCTTTC 120
Db 1612 AAAACCAATGACGATTTTATGTCCTCCGATCGTTTGTCTTATGTTGCTTTC 1671

Oy 121 ATGAGGGGGGAAAAAGCGGAAAGCGGAAAGCTAAAGCTTTCGAGAGAGTACTCTCA 180
Db 1672 ATGAGGGGGGAAAAAGCGGAAAGCGGAAAGCTTTCGAGAGAGTACTCTCA 1731

Oy 181 GCCTCATTTGAGCGGATATTATGATGGAGCGCTATGTTGAAAAAGTAACGAGCAAAA 240
Db 1732 GCCTCATTTGAGCGGATATTATGATGGAGCGCTATGTTGAAAAAGTAACGAGCAAAA 1791

Oy 241 GGGGCGACGATTTTTCATATGTTGAACCGATCTTTAGAGCTCGGATTTATGACA 300
Db 1792 GGGGCGACGATTTTTCATATGTTGAACCGATCTTTAGAGCTCGGATTTATGACA 1851

Oy 301 GGAACCTTTGAAACCCGGTAAACCATCAAAAGATTTTAAACAAGACGATTAAGAGATT 360
Db 1852 GGAACCTTTGAAACCCGGTAAACCATCAAAAGATTTTAAACAAGACGATTAAGAGATT 1911

Oy 361 CATCTGACAGCAATTAAGCAATCAGTGAAGCTCTTGAAGATATGATTTGACGGTTCTC 420
Db 1912 CATCTGACAGCAATTAAGCAATCAGTGAAGCTCTTGAAGATATGATTTGACGGTTCTC 1971

Oy 421 AACAGCGCAACACCGCAATGATTAACGGCTTCAAGGGCATTAAGATACGTTTGA 480
Db 1972 AACAGCGCAACACCGCAATGATTAACGGCTTCAAGGGCATTAAGATACGTTTGA 2031

Oy 481 GAATTTGGAAAGCAAAACCTTGATTCGTTGGAGGGGAGTACAGTTTAAGATGCGGAA 540
Db 2032 GAATTTGGAAAGCAAAACCTTGATTCGTTGGAGGGGAGTACAGTTTAAGATGCGGAA 2091

Oy 541 AAGAAAAATTTGCTACAGAAAGTCAACGGGGTAAACGATTTGACGCTTTACCGAT 600
Db 2092 AAGAAAAATTTGCTACAGAAAGTCAACGGGGTAAACGATTTGACGCTTTACCGAT 2151

Oy 601 GTGTCCGGAAAAAGTTTGGCGGCTTAAAAAGATACGCGGGCGTGTCTCCGCAATCCT 660
Db 2152 GTGTCCGGAAAAAGTTTGGCGGCTTAAAAAGATACGCGGGCGTGTCTCCGCAATCCT 2211

Oy 661 GAAATCTTCATCCCTATGATTTCAAGAGAAAAAATGCTGACATTTGTTGTGAG 720
Db 2212 GAAATCTTCATCCCTATGATTTCAAGAGAAAAAATGCTGACATTTGTTGTGAG 2271

Oy 721 TCACACTGGGCGCAAGAGATGATGACATGATCCAAACGACCGCGCAGCTTCAAGA 780
Db 2272 TCACACTGGGCGCAAGAGATGATGACATGATCCAAATGACCGCGCAGCTTCAAGA 2331

Oy 781 GCCATGCTGATGCGGAGCTGACATCATCGTGGCCATCATCGCAAGCTTTAGAACCG 840
Db 2332 GCCATGCTGATGCGGAGCTGACATCATCGTGGCCATCATCGCAAGCTTTAGAACCG 2391

Oy 841 ATTGAAGATATTAAGGAACCTCTATTTTCTACAGCTTCGCAACTTTGTCTTTACCAA 900
Db 2392 ATTGAAGATATTAAGGAACCTCTATTTTCTACAGCTTCGCAACTTTGTCTTTACCAA 2451

Oy 901 GGCTGAGAGAAACAAGACAGTGCACGTGTTCAAGTATCACTGAAGAAAAATGGAACA 960

```

```

Db 2452 GGCTGACGAGAAACAAGACAGTGCACGTGTTCAAGTATCACTGAAAGAAAAATGGAACA 2511

Oy 961 GGCGGCTTTGAAGTACACCGATGATATCCATGAACGACACCTGACCGTGAAGAAA 1020
Db 2512 GGAGCGCTTTGAAGTACACCGATGATATCCATGAACGACACCTGCGCTGTGAAGAAA 2571

Oy 1021 GACAGCCTTAAACAGAAAAACCATATTTCGCGAACTGACGAAAGACTTAATTTGCTTGG 1080
Db 2572 GACAGCCTTAAACAGAAAAACCATATTTCGCGAACTGACGAAAGACTTAATTTGCTTGG 2631

Oy 1081 AAGTAGAAGACGGAACCTGACCTTGTATTTGATCATATGTCACAAACTTAAATCTTAA 1140
Db 2632 AAGTAGAAGACGGAACCTGACCTTGTATTTGATCATATGTCACAAACTTAAATCTTAA 2691

RESULT 7
BSUB0019/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
REFERENCE
AUTHORS

Bacillus subtilis subsp. subtilis str. 168
Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 200690)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Berteiro, M.G., Bessières, P., Bolo, A., Borchert, S.,
Borries, R., Boursier, L., Brans, A., Braun, M., Brinell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrati, E.,
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaeser, P., Goffeau, A., Golligly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Hage, K., Halech, J., Harwood, C.R.,
Henaud, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koeter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapdus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Maeda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moesli, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S., Parro, V., Pohl, T.M.,
Portecelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynold, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serro, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpe, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassart, A.,
Vial, A., Wambut, R., Wedler, E., Wedler, H., Welter, T.,
Winters, P., Wipac, A., Yamamoto, H., Yamane, K., Yasunoko, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbach, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
9384377
2 (bases 1 to 200690)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur,
Génétique des Génomes Bactériens, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr
Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
On Jul 7, 2003 this sequence version replaced gi:2636029.
This entry contains data from release R16.1 of the Substifiet
database. Further data on gene annotation and detailed information

```


Query Match	98.9%; Score 1127.2; DB 1; Length 200690;
Best Local Similarity	99.3%; Pred. No. 9.3e-276;
Matches 1132; Conservative	0; Mismatches 8; Indels 0; Gaps 0;
Qy	1 ATGAAAAAGACTGAGCTTTCATGAAAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAAG 60
Db	89010 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAAG 88951
Qy	61 AAAACCAATAAGCAGTATTTATGCGCATTCGATCGTTTTGTCCTATGTTGCTTC 120
Db	88950 AAAACCAATAAGCAGTATTTATGCGCATTCGATCGTTTTGTCCTATGTTGCTTC 88891
Qy	121 ATGTGGCGGGAAAAAGCGGAAACGCCGGAAGTCAAAAGTATTCGACGACGCTCTCA 180
Db	88890 ATGTGGCGGGAAAAAGCGGAAACGCCGGAAGTCAAAAGTATTCGACGACGCTCTCA 88831
Qy	181 GCCTCATTTGTAGGCGATTTATGATGGGAGCGCTATGTTGAAAAAGTAAACGAGCAAAAA 240
Db	88830 GCCTCATTTGTAGGCGATTTATGATGGGAGCGCTATGTTGAAAAAGTAAACGAGCAAAAA 88771
Qy	241 GGGGAGACAGTATTTTCAATATGTTGAACGCAATCTTAGAGCCTCGGATTTATGTAGCA 300
Db	88770 GGGGAGACAGTATTTTCAATATGTTGAACGCAATCTTAGAGCCTCGGATTTATGTAGCA 88711
Qy	301 GGAACCTTTGAAAAACCCGTAACCTATCAAAAAGATTATTAACAAGCAGATTAAGAGATT 360
Db	88710 GGAACCTTTGAAAAACCCGTAACCTATCAAAAAGATTATTAACAAGCAGATTAAGAGATT 88651
Qy	361 CATCTGCAGACGAATTAAGAACTCACTGAAAGTCTTGAAAGATATGATTAATTTCAAGTTTC 420
Db	88650 CATCTGCAGACGAATTAAGAACTCACTGAAAGTCTTGAAAGATATGATTAATTTCAAGTTTC 88591
Qy	421 AACAGCGCAACAACACAGCAATGGAATTACGGCGTTTCAGGGGATGAAGATACGCTTGA 480
Db	88590 AACAGCGCAACAACACAGCAATGGAATTACGGCGTTTCAGGGGATGAAGATACGCTTGA 88531
Qy	481 GAATTTGCGAAGCAAAACCTTGATTCGTTGAGCGGAGTACAGCTTAAGTATGATCGGAA 540
Db	88530 GAATTTGCGAAGCAAAACCTTGATTCGTTGAGCGGAGTACAGCTTAAGTATGATCGGAA 88471
Qy	541 AAGAAAATTTGCTACAGAAAGTCAACGGGGTAAACGATTGCAACGCTTGCTTTACCGAT 600
Db	88470 AAGAAAATTTGCTACAGAAAGTCAACGGGGTAAACGATTGCAACGCTTGCTTTACCGAT 88411
Qy	601 GTGTCGGGGAAAAAGTTTGGCGCTAAAAAAGATAGCCGGGGGTGCTCCGCGACATCCT 660
Db	88410 GTGTCGGGGAAAAAGTTTGGCGCTAAAAAAGATAGCCGGGGGTGCTCCGCGACATCCT 88351
Qy	661 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGACATTTGTTGTGCGAG 720
Db	88350 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGACATTTGTTGTGCGAG 88291
Qy	721 TCACACTGGGGCCAAAGATATGACATATCCAAAGACCGCGACGCCAGCTTGCAAGA 780
Db	88290 TCACACTGGGGCCAAAGATATGACATATCCAAAGACCGCGACGCCAGCTTGCAAGA 88231
Qy	781 GCCAATGCTGATGGGGAGCTGACATCATCGTGGCGCATCATCCGCAAGCTCTTAGAACG 840
Db	88230 GCCAATGCTGATGGGGAGCTGACATCATCGTGGCGCATCATCCGCAAGCTCTTAGAACG 88171
Qy	841 ATTGAAGTATATTAACGGAACCGTCATTTTCTACAGCCTCGGCAACTTTGTCTTTGACCA 900
Db	88170 ATTGAAGTATATTAACGGAACCGTCATTTTCTACAGCCTCGGCAACTTTGTCTTTGACCA 88111
Qy	901 GGCTGAGCAGAGACAAGACAGTGCATCTGTTCAAGTATCACTGAAAGAAAAATGAAACA 960
Db	88110 GGCTGAGCAGAGACAAGACAGTGCATCTGTTCAAGTATCACTGAAAGAAAAATGAAACA 88051
Qy	961 GGCGCGCTTTGAAGTGACACCGGATCGATATCCATGAAAGCAGCCCTGCTGAAAAAA 1020
Db	88050 GGCGCGCTTTGAAGTGACACCGGATCGATATCCATGAAAGCAGCCCTGCTGAAAAAA 87991
.Qy	1021 GACAGCCTTTAAACAGAAAAACATATTGCGGAAGTACGAAAGACTTAATTTCCGTTGG 1080

Db	87990 GACAGCCTTTAAACAGAAAAACCATATTTCGGAACTGACGAAAGACTCTTAATTTGCTTGG 87931
Qy	1081 AAAGTAGAGCGGAAAAAGCTGAGTTGATATGATCATATGATGACAAACTAAATCTTAA 1140
Db	87930 AAAGTAGAGCGGAAAAAGCTGAGTTGATATGATCATATGATGACAAACTAAATCTTAA 87871
RESULT 8	
LOCUS	DQ086153 2775 bp DNA linear BCT 27-JUN-2005
DEFINITION	Bacillus subtilis strain ZJU-7 Pgab (pgab), pgac (pgac), and PgaA
ACCESSION	DQ086153
VERSION	DQ086153.1 GI:68138277
KEYWORDS	
SOURCE	Bacillus subtilis
ORGANISM	Bacillus subtilis
REFERENCE	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
AUTHORS	Shi, F., Xu, Z., and Cen, P.
TITLE	Efficient production of poly(glutamic acid) by a new strain
JOURNAL	Bacillus subtilis ZJU-7
REFERENCE	2 (bases 1 to 2775)
AUTHORS	Shi, F., Xu, Z., and Cen, P.
TITLE	Direct Submission
JOURNAL	Submitted (04-JUN-2005) Institute of Bioengineering, Department of Chemical Engineering and Bioengineering, Zhejiang University, No.38, Zheda Road, Hangzhou, Zhejiang 310027, P.R. China
FEATURES	Location/Qualifiers
source	1..2775
	/organism="Bacillus subtilis"
	/mol_type="Genomic DNA"
	/strain="ZJU-7"
	/db_xref="taxon:1423"
gene	1..1182
	/gene="pgab"
CDS	1..1182
	/function="biosynthesis of poly(glutamic acid)"
	/codon_start=1
	/transl_table=1
	/product="Pgab"
	/protein_id="AAV85940.1"
	/db_xref="GI:68138278"
gene	1183..1632
	/gene="pgac"
CDS	1183..1632
	/function="biosynthesis of poly(glutamic acid)"
	/codon_start=1
	/transl_table=1
	/product="Pgac"
	/protein_id="AAV85941.1"
	/db_xref="GI:68138279"
translation="MFSSDYIALILVLSLIFAEKTIQVPAQVYVPGVLGVNPD	
VFILLVLSLIVLVIVKYGSKRMILYGRKRKPMILITGVLKIAFDPLPIVFEI	
AEFGIIGIYVGLANTIKQGLITITGSLTLLSGATFALMFYLLI"	
1633..2775	
gene	/gene="pgaa"
CDS	1633..2775
	/gene="pgaa"
	/function="biosynthesis of poly(glutamic acid)"
	/codon_start=1

```

/translate=1
/product="Pga"
/protein_id="AAY85942.1"
/db_xref="GI:68138280"
/translation="MKKELSPHEKLLKTKOOKKKKKNHVEIALIPVLFVAFEMWAG
KAEPKVTYDDVLSASFVGDIMRGVETBEKGDLSILOYERFERASDYVAGN
FENPTTYKRYKQADKEHLOTKNESVULADMDNPTVLNANNHMDGVGMDOTLG
EPKQNDLIVGAGKSLSDAKKISLYOKNGVTIATLGTIDVSGRPAKKNTPGTLPA
DPELIPMISEKKHADIIVVQSHWGOBYDNDROKOLAKMAGSDAGDILIVGHHPH
VLEPIEVNGTIVIFYSIGNFVFDQMTIRDSALVOYHLKNGTGRFEVPTDIHEAT
PAPVKDLSIKQITIRLELTKDSNFAMKVEDKLFEDIDHSDLAKSK"

```

ORIGIN

```

Query Match      98.5%; Score 1122.4; DB 1; Length 2775;
Best Local Similarity 99.0%; Pred. No. 1.3e-274;
Matches 1129; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGAAAAAGAACTAGCTTTTCATGAAAAAGCTGAAGCTGACAAACAGCAAAAAAG 60
DB 1633 ATGAAAAAGAACTAGCTTTTCATGAAAAAGCTGAAGCTGACAAACAGCAAAAAAG 1692
QY 61 AAAACCAATAGCAGTATTTTATGCAATCCGATCGTTTGTCTTATGTTGCTTTC 120
DB 1633 AAAACCAATAGCAGTATTTTATGCAATCCGATCGTTTGTCTTATGTTGCTTTC 1752
QY 121 ATGTGGCGGGGAAAAAGCGGAAAGCTCAAAAGCTATTCTGACGACTCTTCA 180
DB 1753 ATGTGGCGGGGAAAAAGCGGAAAGCTCAAAAGCTATTCTGACGACTCTTCA 1812
QY 181 GCCTCATTTGTAGGGGATTTATGAGTGGGACGCTATGTTGAAAAAGTAACGAGCAAAA 240
DB 1813 GCCTCATTTGTAGGGGATTTATGAGTGGGACGCTATGTTGAAAAAGTAACGAGCAAAA 1872
QY 241 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGACCTCGGATTAATGAGA 300
DB 1873 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGACCTCGGATTAATGAGA 1932
QY 301 GGAACCTTTGAAACCCGGTAACCTATCAAAAGATTTAACAACAGCAATAAGAGATT 360
DB 1933 GGAACCTTTGAAACCCGGTAACCTATCAAAAGATTTAACAACAGCAATAAGAGATT 1992
QY 361 CATTCGACAGCAATTAAGATATGTAAGTCTTGAAGATTAATTCACGGTTTC 420
DB 1993 CATTCGACAGCAATTAAGATATGTAAGTCTTGAAGATTAATTCACGGTTTC 2052
QY 421 AACAGCGCAACCAACGCAATGATTCAGCGGTCAGGGGCAATGAAGTACGTTGGA 480
DB 2053 AACAGCGCAACCAACGCAATGATTCAGCGGTCAGGGGCAATGAAGTACGTTGGA 2112
QY 481 GAATTCGCAAGCAAAACCTGATATCGTTGAGCGGATACAGCTTAAGTAGTGCAGA 540
DB 2113 GAATTCGCAAGCAAAACCTGATATCGTTGAGCGGATACAGCTTAAGTAGTGCAGA 2172
QY 541 AAGAAAAATTTCTGACGAAGATCAACGGGGTAAAGATTGCAACGTTGGCTTTACCGAT 600
DB 2173 AAGAAAAATTTCTGACGAAGATCAACGGGGTAAAGATTGCAACGTTGGCTTTACCGAT 2232
QY 601 GTGTCCGGGAAAGTTTCGGCGCTAAAGAAATAGCGCGGGCGTGCACCGAGATCTT 660
DB 2233 GTGTCCGGGAAAGTTTCGGCGCTAAAGAAATAGCGCGGGCGTGCACCGAGATCTT 2292
QY 661 GAAATCTTCACTCCCTATGATTTGAGAGCGAAAAAATCATGCTGATGTTGTTGTCAG 720
DB 2293 GAAATCTTCACTCCCTATGATTTGAGAGCGAAAAAATCATGCTGATGTTGTTGTCAG 2352
QY 721 TCACACTGGGGCCAAAGATATGACATATGTCAAACGACCGCCACGCTTTCAGAGA 780
DB 2353 TCACACTGGGGCCAAAGATATGACATATGTCAAACGACCGCCACGCTTTCAGAGA 2412
QY 781 GCCATGCTGATGGGAGCTGACATCATGCTGGCCATCACCGCAGCTTTAGAACCG 840
DB 2413 GCCATGCTGATGGGAGCTGACATCATGCTGGCCATCACCGCAGCTTTAGAACCG 2472

```

```

QY 841 ATTGAGATATATAACGAAACCGTCAATTTTTCATACAGCTCGGCAACTTTGTCTTGACCA 900
DB 2473 ATTGAGATATATAACGAAACCGTCAATTTTTCATACAGCTCGGCAACTTTGTCTTGACCA 2532
QY 901 GGTGGACGAGAACAGACAGACAGTGCATGTTCAATACCTGAGAAAAATGGAACA 960
DB 2533 GGTGGACGAGAACAGACAGACAGTGCATGTTCAATACCTGAGAAAAATGGAACA 2592
QY 961 GGGCGCTTTGAAGTACGACCGGATGATTCATGAAAGGACACCTGCACCTGTGAAAAAA 1020
DB 2593 GGGCGCTTTGAAGTACGACCGGATGATTCATGAAAGGACACCTGCACCTGTGAAAAAA 2652
QY 1021 GACAGCCTTAAACAGAAACCAATTAATTCGCAACTGACGAAAGACTTAATTTGCTTGG 1080
DB 2653 GACAGCCTTAAACAGAAACCAATTAATTCGCAACTGACGAAAGACTTAATTTGCTTGG 2712
QY 1081 AAAGTAGAAGACGGAACCTGACGTTGATATGATCACTAGTACAAACTAAATCTTAA 1140
DB 2713 AAAGTAGAAGACGGAACCTGACGTTGATATGATCACTAGTACAAACTAAATCTTAA 2772

RESULT 9
AE017333 36/c
WPCOMMENT
Sequence split into 43 fragments LOCUS AE017333 Accession AE017333
Fragment Name Begin End
AE017333_00 1 110000
AE017333_01 100001 210000
AE017333_02 200001 310000
AE017333_03 300001 410000
AE017333_04 400001 510000
AE017333_05 500001 610000
AE017333_06 600001 710000
AE017333_07 700001 810000
AE017333_08 800001 910000
AE017333_09 900001 1010000
AE017333_10 1000001 1110000
AE017333_11 1100001 1210000
AE017333_12 1200001 1310000
AE017333_13 1300001 1410000
AE017333_14 1400001 1510000
AE017333_15 1500001 1610000
AE017333_16 1600001 1710000
AE017333_17 1700001 1810000
AE017333_18 1800001 1910000
AE017333_19 1900001 2010000
AE017333_20 2000001 2110000
AE017333_21 2100001 2210000
AE017333_22 2200001 2310000
AE017333_23 2300001 2410000
AE017333_24 2400001 2510000
AE017333_25 2500001 2610000
AE017333_26 2600001 2710000
AE017333_27 2700001 2810000
AE017333_28 2800001 2910000
AE017333_29 2900001 3010000
AE017333_30 3000001 3110000
AE017333_31 3100001 3210000
AE017333_32 3200001 3310000
AE017333_33 3300001 3410000
AE017333_34 3400001 3510000
AE017333_35 3500001 3610000
AE017333_36 3600001 3710000
AE017333_37 3700001 3810000
AE017333_38 3800001 3910000
AE017333_39 3900001 4010000
AE017333_40 4000001 4110000
AE017333_41 4100001 4210000
AE017333_42 4200001 4222645
Continuation (37 of 43) of AE017333 from base 3600001 (AE017333 Bacillus licheniformis D5
Query Match 47.4%; Score 540.2; DB 1; Length 110000;
Best Local Similarity 68.2%; Pred. No. 2.7e-126;

```

Matches 781; Conservative 0; Mismatches 358; Indels 6; Gaps 2;

OY 1 ATGAAAAAGAACTGAGCTTTTCATGAAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
DB 65919 ATGAAAAAACAAGTAACTTTTCAGAAAAAAGCTGCAAGTTGACCAACAGAGAAAAAG 65860
OY 61 AAAACCAATAGACAGTATTTATTTCCATTCGATCGTTTGTCTTATGTCGCTTC 120
DB 65859 AAAACCAACAAAGCAGCTCTTTATCTGATGCCGTTATTTCTGTATTAATGTTGTCTT 65800
OY 121 ATGTGGCGGGAAGGCGAAACGC--CGAAGTCAAAAGTATTGACGACGTAATC 177
DB 65799 ACTTGGGTGGAGGCGCAAACTCTTCGCAAAATGCAAAAAAGAAAGTCCAACTT 65740
OY 178 TCAGGCTATTGTAGGAGATTTATGATGGGACGCTATGTTGAAAAAGTAAAGGAGCA 237
DB 65739 ACAGTACTTTTGTGGGATATCATGATGGAGAAACGTAAGAAAAAGTGAACAACTTG 65680
OY 238 AAAGGGGAGACAGTATTTTCAATATGTTGAAACCGATCTTTAGAGCTCGGATATGTA 297
DB 65679 CACGTTTGGAAGGTCTTCAAAAATGTAAGCCGTACTTTAATGTCTCAGATTTTATC 65620
OY 298 GCAGGAACCTTTGAAACCCGGTAACTATCAAAAGAAATTAATTAACAAGCAGATTAAG 357
DB 65619 ACAGGAACCTTTGAAACCCGTAACTATCAAAAGAAATTAATTAACAAGCAGATTAAG 65560
OY 358 ATTATCTGCAACCAATTAAGAAATCAGTGAAGCTTTGAAAGATTAATTTCAAGCTT 417
DB 65559 ATCAATCTGCAACCAATTAAGAAATCAGTGAAGCTTTGAAAGATTAATTTCAAGCTT 65500
OY 418 CTCACAGCGCCAAACCAACGAGTATTCGCGCTTTCAGGGGATGAAAGATACGCTT 477
DB 65499 CTGAATTTTGGCAACACATGCGATGACATCGGGAAGACGTTTGAAGATACGCTC 65440
OY 478 GGAGATTTGCGAGCAAAACCTTGATATCTGTGAGCGGGATACAGCTTAAGTATCG 537
DB 65439 AATAAATTTTCAATGAGAAATCTGAGCTTGTGCGAGCAAGAAATTAATTTGAAACCG 65380
OY 538 AAAAAGAAATTTGCTACCAAGAAATCAACGGGATTAAGATGCAACCTTGGCTTAC 597
DB 65379 AAACGACGATCTTATCAAGATGTAACGGCGTAAATTTGCAACGCTCGTTTACA 65320
OY 598 GATGTGTCGGGAAAGTTTCGCGCTTAAAGAAATACGCCGGGCTGTCGCCGACAT 657
DB 65319 GACGTCTACAAAGAACCTTTACACCAAAAAGAACAGGGGAGTGTCTGCTCAGT 65260
OY 658 CCGTGAATCTTCAATCTTATGATTTTCAAGAGCGAAAAAATGCTGACATGTTGTGTG 717
DB 65259 CCGAAATCTTATCTCAATGATTTGCGAAGCATGCAAAAAAGCGAATCTTGTCTTGTG 65200
OY 718 CAGTCACTGGGGCAAGAGATTAAGTAATGATCCAAAGACCGGACGCAAGTTCGA 777
DB 65199 CATGTGCACTGGGCAAGAAATATGACAAATGAAAGCAAGACAGAAAGATGTGCC 65140
OY 778 AGAGCAGTCTGATGCGGAGCTGACATCATCTGCGCATCTCCGACGCTTTTAA 837
DB 65139 AAGGCGATTCAGATGCGGAGCAGATGTCATCATCGGCGCTCATCCCATGTTCTCAA 65080
OY 838 CCGATTAAGTATATACGGAACCGTCAATTTTCAACGCTTCGGCACTTGTCTTGA 897
DB 65079 CCGATCGAAGTGTATACGGTACTGTGATTTTCTACAGCTTCGCAACTTTGTATTTAT 65020
OY 898 CAAGCTTGACGAAACAAAGACAGTGCACGTGTTCAATCATCTGAGAAAAATGA 957
DB 65019 CAGGCTGTGTCAAGAACAGGAGACGCGCTTGTACAAATCAATTTATGATGATGACGC 64960
OY 958 ACAGGCGCTTTGAAGTGACACGATCGATTCATTCATGAGCAGACCTGCACT--GTG 1014
DB 64959 AAAGGCGCTTTGAGATACGCTCTCAACATTCGGAAGCAGACCCGACGCTTTTATG 64900
OY 1015 AAAAAAGACAGCTTTAAACAGAAACCAATATTCGCAACTGACGAAAGATCTTAATTC 1074
DB 64899 AAGAGCAGCTTCTTAAAGCAAAAGCAATCTTCGCTCAATTTGACAAAGAAACAACTC 64840

OY 1075 GCTTGAAGTGAAGACGGAAGAACTGACGTTTGTATATGATCATAGTCAAACTAAAA 1134
DB 64839 GACTGGAAGAGAGAACGGAATTAACGTTTGAATGATCATGATGCGACAACTGAAA 64780
OY 1135 TCTAA 1139
DB 64779 AATAA 64775

RESULT 10
CP000002_36/c
WPCOMMENT
Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000
CP000002_11	1100001	1210000
CP000002_12	1200001	1310000
CP000002_13	1300001	1410000
CP000002_14	1400001	1510000
CP000002_15	1500001	1610000
CP000002_16	1600001	1710000
CP000002_17	1700001	1810000
CP000002_18	1800001	1910000
CP000002_19	1900001	2010000
CP000002_20	2000001	2110000
CP000002_21	2100001	2210000
CP000002_22	2200001	2310000
CP000002_23	2300001	2410000
CP000002_24	2400001	2510000
CP000002_25	2500001	2610000
CP000002_26	2600001	2710000
CP000002_27	2700001	2810000
CP000002_28	2800001	2910000
CP000002_29	2900001	3010000
CP000002_30	3000001	3110000
CP000002_31	3100001	3210000
CP000002_32	3200001	3310000
CP000002_33	3300001	3410000
CP000002_34	3400001	3510000
CP000002_35	3500001	3610000
CP000002_36	3600001	3710000
CP000002_37	3700001	3810000
CP000002_38	3800001	3910000
CP000002_39	3900001	4010000
CP000002_40	4000001	4110000
CP000002_41	4100001	4210000
CP000002_42	4200001	422234

Continuation (37 of 43) of CP000002 from base 3600001 (CP000002 Bacillus licheniformis AT)

Query Match 47.4%; Score 540.2; DB 1; Length 110000;
Best Local Similarity 68.2%; Pred. No. 2.7e-126;
Matches 781; Conservative 0; Mismatches 358; Indels 6; Gaps 2;

OY 1 ATGAAAAAGAACTGAGCTTTTCATGAAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
DB 65759 ATGAAAAAACAAGTAACTTTTCAGAAAAAAGCTGCAAGTTGACGAAAGCAAGAAAAAG 65700
OY 61 AAAACCAATAGACAGTATTTATTTCCATTCGATCGTTTGTCTTATGTCGCTTC 120
DB 65699 AAAACCAACAAAGCAGCTCTTTATCTGATGCCGTTATTTCTGTATTAATGTTGTCTT 65640
OY 121 ATGTGGCGGGAAGGCGAAACGC--CGAAGTCAAAAGTATTGACGACGTAATC 177

```

Db 65639 ACTGGGCGAAGGCGCAAACTCTTCGCAATGCAAAAAAGAGTCCAACTT 65580
Oy 178 TCAGCCTCATTTGAGCGATATTATGATGGACCGCTATGTTGAAAAAGTAAACGAGAGA 237
Db 65579 ACAGTACTTTTGTGGGATATCATATGATGGAAAGAAAGTAAAGTGAACAACTTG 65520
Oy 238 AAAGGGGAGACAGTATTTTCAATATGTTGAACCGATCTTTAGACCTCGATTAATGA 297
Db 65519 CACGGTTCGAAAGTGTCTTCAAAAATGTGAAGCCGTCTTTAAATGTGTCAATTTATC 65460
Oy 298 GCAGAACTTTGAAACCCGCTACCTATCAAAAGATTTAAACAAGCAGATAAAG 357
Db 65459 ACAGAACTTTGAAACCCGCTATCAATGCAAGAGCATTCAGAGCGCAAGAAAGAC 65400
Oy 358 ATTCACTGACAGCAATATAGGATCAGTGAAGTCTTGAAAGATATGATTTCAACGTT 417
Db 65339 ATCCATCTGCAAAACCAATCAAGATCAGTCAAAATTTGAAAGCTGAACCTGACGCTA 65340
Oy 418 CTCACAGCGCCACCAACCAACGCAATGATTAACGCGCTTCAGGCAATGAAGATACGCTT 477
Db 65339 CTGAATTTTGGCCCAACCAACCATGCGATGACCTACGCGGAAAGACGTTTGAAGATACGCTC 65280
Oy 478 GGAAGATTTGGCAACCAAACTTGTATATGTTGAGCGGGAATACAGCTTAAGTATGCG 537
Db 65279 AATMAATTTTCAAAATGAGATGTGAGCTTGTGCGAGCGAAATATCTTGAAAGCGCG 65220
Oy 538 AAAAGAAATTTTCGACAGAAAGTCAACGCGGATTAACGATTTGCAACGCTTGTTACC 597
Db 65219 AAACGACAGTATCTTATCAATGATGAGCGCGTAAATTTGCAACGCTCGGTTTACA 65160
Oy 598 GATGTGTCGCGGAAAGGTTTCGCGCTAAAGAAATACGCGCGGCTGTGCTGCCGACAT 657
Db 65159 GACGTCTACAAAGAACTTTACAGCAAAAGAAACAGAGCGGAGTGTGCTGCCGCTACAT 65100
Oy 658 CCGTAATTTTCAATCCCTTATGATTTTCAAGAGGAAAAACATGCTGACATTTGTGTG 717
Db 65099 CGAAATATCTTTATTCATGATGATTTGGGAGAGCATCGAAAAAGCGATCTTGTCTTGTG 65040
Oy 718 CAGTCACTGGGGCGCAAGATGATGACATGATTCGAAAGCAACGCGCAACGCGCACTTGA 777
Db 65039 CATGTGCACTGGGGCGCAAGATGATGACATGATTCGAAAGCAACGCGCAACGCGCACTTGA 64980
Oy 778 AGAGCATGTCTGATGCGGAGCTGACATCATGCTGCGCATCATTCGCAACGCTTTAGA 837
Db 64979 AAGGGATTTGAGATGCGGAGCTGACATCATGCTGCGCATCATTCGCAACGCTTTAGA 64920
Oy 838 CCGATTTGAAGTATTAACGAAACCGTCAATTTTCAAGCTTGGGCACTTTGTCTTGAC 897
Db 64919 CCGATCGAAGTGTATTAACGATCTGTGATTTTCTACAGCTCGGCAACTTTGTATTTGAT 64860
Oy 898 CAAGGTGAGCAAGCAAGAGACAGTGAACGCTGACATGATCACTGAAGAAAAATGA 957
Db 64859 CAGGGTGTCTAAGAAACAGGAGCAGCGGCTTTGTAACATCAATTTAAATGAATGACGGC 64800
Oy 958 ACAGCGCGCTTGAAGTGAACCGATCGATGATTCATGAAAGCGCAACCTGACCT--GTG 1014
Db 64799 AAAGGCGCTTTGAGTAAAGCTCTCAACATTCGGAAGCAACGCGCAACCTTTAGGC 64740
Oy 1015 AAAAAAGACAGCTTTAAACAGAAACCATTAATTCGGAAGTGAAGCAAGACTTAATTTG 1074
Db 64739 AAGAGGAGCTTTCTTAACAGAAAGGATCTTCCGTAATTTGAACAAAGGAAACAACTTC 64680
Oy 1075 GCTTGAAGTGAAGCAAGCAAGCAAGCAAGCTTGAATGATGATGATGATGATGATGATGAT 1134
Db 64679 GACTTGAAGCAAGCAAGCAAGCAAGCAAGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 64620
Oy 1135 TCTAA 1139
Db 64619 AATTA 64615

```

```

AX433330
LOCUS AX433330 696 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 1745 from Patent WO0229113.
ACCESSION AX433330
VERSION AX433330.1 GI:21658134
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1. Berka, R. and Clausen, I. G.
AUTHORS Methods for monitoring multiple gene expression
TITLE Patent: WO 0229113-A 1745 11-APR-2002;
JOURNAL Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
source
1. .696
/organism="Bacillus licheniformis"
/mol_type="unassigned DNA"
/db_xref="taxon:1402"
ORIGIN
Query Match 29.2%; Score 332.4; DB 6; Length 696;
Beet Local Similarity 68.4%; Pred. No. 1.8e-73;
Matches 475; Conservative 0; Mismatches 216; Indels 3; Gaps 1;
Oy 444 GGATTACGGCGCTTCAGGCGCATGAAAGATACGCTTGGAGATTTGGAGCAAAACCTTGA 503
Db 1 GACTACCGGGAAGCGGTTTGAAGATACGCTCAATTAATTTTCAATGAGATCTGA 60
Oy 504 TATCGTTGAGCGGGATACAGCTTAAGTATGATCGCAAAAGAAATTTGTAACAGAAAGT 563
Db 61 GCTTTCGAGACAGAAATTAATCTTGAAGACGCGAAACAGACGATCTCTATCAGAAATG 120
Oy 564 CAACGGGTTACGATTCAGCGCTTTCAGCTTATCCATGTGTCCGGAAGAGTTTCCGCGC 623
Db 121 GAACGGCGTAAATTAATGCAACGCTCGTTTACAGACGCTTACACAAAGAACTTTACAGC 180
Oy 624 TAAAGAAATACCGCGCGCTGCTGCGCGAGATCTCTGAATCTCATCCATGATTTG 683
Db 181 CAAAAAGAAACAGAGCGGAGTGTCTCCCTCAGTCCGAAATCTTTATTCATGATTTGC 240
Oy 684 AGAGCGAAAAACATGCTGACATTTGTTGTGTGACATCAGCTGCGGCGCAAGATGTA 743
Db 241 GAAAGATCGAAAAAGCGGATCTGTCTCTTCCATGATGCACTGGGCAAGAAATATGA 300
Oy 744 CAATGATCCAAACGACCGCGACGCGCAGTTGCAAGAGCATGTCTGATGCGGAGCTGA 803
Db 301 CAATGAACGAAACGACAGCAGAAAGATCTGGCAAGCGATGCAATGCGGAGCAGA 360
Oy 804 CATCATGCTGCGCATCATCCGCAAGTCTTAAGAACCGATTAATTAAGGAAACGCT 863
Db 361 TGTCAATCATGCGCGCTCATCTCCCATGTCTCTCAACCGATGCAAGTGTATTAAGGTA 420
Oy 864 CATTTTCTACAGCTCGGCACTTTGTCTTTGACCAAGGCTGAGCAAGAAAGAGACAG 923
Db 421 GATTTTCTACAGCTCGGCACTTTGTATTTATATAGGCTGTGTAAGAAACAGGAGAG 480
Oy 924 TCACTGTTCAATATCACTGCAAGAAATGAAACAGCGCTTTGAAAGTGAACCGAT 983
Db 481 CCGCTGTGCAATATCAATTAATGATGACGCAAGGCGCTTTGAGTAAACGCTCT 540
Oy 984 CGATATCATGAAAGCAACCTGCACT--GTGAAAAAGCAGCTTTAAACAGAAAC 1040
Db 541 CAACATTCGCGAAGCAACCGCGCTTTAGCAAGACGACTTTAAACGAAAGAC 600
Oy 1041 CATTAATGCGCAAGCAAGCAAGCAAGCTCTAATTTGCTTGAAGTGAAGCAAGCAAGCT 1100
Db 601 GATCTTCGCTCAATTAAGCAAGCAAGCAAGCTCTGACACTGAAAGAAAGCAAGCAAGCT 660
Oy 1101 GACGTTGATATGATCATATGACAAACTAAA 1134
Db 661 AACGTTGAAGTGAATCATGCGGACAAAGCTGAA 694

```


RESULT 12
AB011191/c 94829 bp DNA circular BCT 14-JUN-2002
LOCUS Bacillus anthracis str. A2012 plasmid pXO2, complete sequence.
DEFINITION AE011191
ACCESSION AE011191
VERSION AE011191.1 GI:20520280
KEYWORDS
SOURCE Bacillus anthracis str. A2012
ORGANISM Bacillus anthracis str. A2012
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
REFERENCE 1 (bases 1 to 94829)
AUTHORS Read,T.D., Salzberg,S.L., Pop,M., Shumway,M., Umayam,L., Jiang,L., Holtzapple,E., Busch,J.D., Smith,K.L., Schupp,J.M., Solomon,D., Keim,P. and Frazer,C.M.
TITLE Comparative genome sequencing for discovery of novel polymorphisms in Bacillus anthracis
JOURNAL Science 296 (5575), 2028-2033 (2002)
PUBMED 12004073
2 (bases 1 to 94829)
Read,T.D., Salzberg,S.L., Pop,M., Shumway,M., Umayam,L., Jiang,L., Holtzapple,E., Busch,J.D., Smith,K.L., Schupp,J.M., Solomon,D., Keim,P. and Frazer,C.M.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 94829
/organism="Bacillus anthracis str. A2012"
/mol_type="genomic DNA"
/strain="A2012"
/db_xref="taxon:191218"
/plasmid="pXO2"
/note="Florida strain"
complement (543..1139)
/gene="BXB0001"
complement (543..1139)
/gene="BXB0001"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pXO2-03)"
/protein_id="AAM26163.1"
/db_xref="GI:20520281"
/translation="MTYVKLFQFRGNQIHVDVDFQRFYEPYREADVLRHAIYVLD EKMBEGKQDDVSLKKEVKDKQSVMTMQQMDVLKLNMLIAQSGSNLSLEKLEQ EVKSEIASVTKSEKSGEPRTVKREPIEENQEHQRIQREVKTSFVKKPLMITDDV QIWKKGKPHQVPTFRIGTGYOEIEMEQIPEHRKLE"
complement (1151..1438)
/gene="BXB0002"
complement (1151..1438)
/gene="BXB0002"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pXO2-04)"
/protein_id="AAM26163.1"
/db_xref="GI:20520282"
/translation="MEIHVRNVDPYHLKEIDRKCEIGKKLGRYYRWEYINMMFEQH FNGEYSRKEDKPEDAVNSITLDRQSDKLQEVYIDATHELVAMIKLNEE"
complement (1652..1768)
/gene="BXB0003"
complement (1652..1768)
/gene="BXB0003"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM26164.1"
/db_xref="GI:20520283"
/translation="MKRKKRYKVFRRIQSAVYRHQNAVYEVLMKNYKAHTI"
complement (1746..2594)
gene

CDS
/gene="BXB0004"
complement (1746..2594)
/gene="BXB0004"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pXO2-05)"
/protein_id="AAM26165.1"
/db_xref="GI:20520284"
/translation="MSIKPWNKEGRKANTKRAYLFTLGSPFTWPFVLCISPVSG ATRKFEWKIGETQSUSSTYKIVAKKEVPPDQTLRIDELRANDSQLSMKRY ENRYIKQKQNNVKTXYRASDNVIVISENVPEEFGVSSVVPETHPBLQNDVLD KERSMKRYVLENELINRELKKSQFYERETLAFSQALRKEIEKMEKSSAMKQD KIKNEOLTKMEYQTEGEKTKNTINSNESTINNHQEIYDKEDIKMEKKIQQLD EKKKT"
complement (2591..2866)
/gene="BXB0005"
complement (2591..2866)
/gene="BXB0005"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pXO2-06)"
/protein_id="AAM26166.1"
/db_xref="GI:20520285"
/translation="MLISELKPNNHDYAKGKYLLSLRKKKGVKDKFIEIPITWPDY NPGKRVWLVIREYQSPVNGEKRYTNCKLENHIAQVSVNVKGERMK"
complement (2955..3563)
/gene="BXB0006"
complement (2955..3563)
/gene="BXB0006"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pXO2-07)"
/protein_id="AAM26167.1"
/db_xref="GI:20520286"
/translation="MNNLSKILSGALVLSFNGIWAYNTTKEDFKDKKRYRVTSK EKOIGDLKVLBOKDKINEGAGENKEENSTLDLONKYREVANQFVHLYDVSQV N KGERRNVLKITDKKYVDIYAPNTDDGDPNFSHYNKAIYINSEBDVSKKCTALLD IETIYELGNKQTTINSVKITLKEQEEIKVAYBNYPVKR"
complement (3589..4731)
/gene="BXB0007"
complement (3589..4731)
/gene="BXB0007"
/note="Identified by match to Pfam protein family HMM PF00877"
/codon_start=1
/transl_table=1
/product="NLP/P60 family protein, (pXO2-08)"
/protein_id="AAM26168.1"
/db_xref="GI:20520287"
/translation="MAITQSTALDKYKVKRIYVLLGSGTGVVIAAITLLIYS MAIFGQSGSTGTPNGGISGTATYKNLPEPEVWRQAWVEQCAAGVELVPYVLAITM VESNGISEKLPIDIMQSSQSGQMANNTISNPKDSIYGVNMLKGFDPARKLGINDLA IVOTYNGRNVVHMLAANKTKHSIQADYVSLTVAPAGGNRNGTTIGSQPVAVANA GGYRIVNGNFEVAEMVKOYLSFPDAGTSGQIDPGSETPKMMDEVLKYNNGNYWVG GKSSSQGPDSGLTYMAKYNAGTIPISAAITQVDFTVENDPKAQRGDLVFFRQTYGG PNHVSHVGIYIDANTMTDNGSGVGTHQFTSSVQGHYAGIRRVPR"
complement (4733..6667)
/gene="BXB0008"
complement (4733..6667)
/gene="BXB0008"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pXO2-09)"
/protein_id="AAM26169.1"
/db_xref="GI:20520288"
/translation="MVKLGKKEQKRYRTBGYDLDFTSRVQPGGIRKNEKYIAGDCY VACHVYSLAEDIPPLWLTALMNNKDTISSVDVATAKKEVVKQINSITELKQRTS ERSTDDDDHWELONTLPDARSITQGEIVKLVKRIYIYDPLVLEKRIQDIKKE IAGQNYAQQYVFKQKEWQTLFASYPDQILEYIGVSGYPLPSKNIGFGIPFHQDLK

DPGVIGTSTGCAFLIDPPSTSTRTSFGFIKGMGAGSKTLLKOLEGLVAKDC
 FIRGDKRDRYTTVVQOQCGKIDISQGDSEHVEKMLNPLVPAATKITTGANDKX
 GSFOHISKVTIMLIFLTEREDVIOERKHLVIFYEMGLVPOQSGDRPKYVIGP
 PNHYILBDFPKYKLSNLEAGATPQRHDLALQVEDIMTVYGDMEHNTLLKNF
 ENBOJLFDIDIGISKYDKSVEFCOLFTKRTIDLSNODLXKRPANDJSLCYVH
 YEMQVLDFCHVINSQNSPANYVLEFREKMKPAGLFPATOSPOEMLPFGSGVDV
 STMKAFELTQXKIFLNMDSVLETKSVLGSLESEFRILPELKRGBALVOVSTE
 TYNVMPDEDEKLERFKGO"
 complement (6684, .7358)
 /gene="BXB009"
 complement (6684, .7358)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein, (px02-10)"
 /protein_id="AA026170.1"
 /db_xref="GI:20520289"
 /translation="MKMKVATPAKKVNRKROISEKSGKOSFISTIFASPKONTAS
 IVDIAPYKLEOKTGLVDHRNNLOVYKTKTIDLSNODLXKRPANDJSLCYVH
 BPFKLSLTVSTETEOOYVWRMALRFQGRMSQVSEKESHELMYQRTSLAENLR
 VLAWEKNLKEALFPIVYVGNKNETELIKVYKDMKRYGQRFNLQNMKAKEVELIFKLQ
 NNSSEM"
 complement (7416, .7730)
 /gene="BXB010"
 complement (7416, .7730)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein, (px02-11)"
 /protein_id="AA026171.1"
 /db_xref="GI:20520290"
 /translation="MEMMNPKNYKQKQIKLAFYIIDGAIILMLVLASVMEKVPVVG
 GGRIMFYVLFGTFGLFCIKXPNSTNNIFVILDMLKMDNKNYHPIEVNTISSEK
 RK"
 complement (7733, .7936)
 /gene="BXB011"
 complement (7733, .7936)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein, (px02-12)"
 /protein_id="AA026172.1"
 /db_xref="GI:20520291"
 /translation="MNNMIFAILMLGVALVLSIAIFPFQIPIFPRKKKEPVWTV
 TNAEYIDRDPFKKKEQKDKKEH"
 complement (7938, .8192)
 /gene="BXB0012"
 complement (7938, .8192)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein, (px02-13)"
 /protein_id="AA026173.1"
 /db_xref="GI:20520292"

Query Match 22.0%; Score 250.4; DB 1; Length 94829;
 Best Local Similarity 55.6%; Pred. No. 1,6e-52;
 Matches 576; Conservative 0; Mismatches 436; Indels 24; Gaps 4;

gene
 CDS
 1 ATGAAAAAGAACTGAGCTTTCATGAAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
 55123 ATGAGAGAAATTTGACATTTTCAAGAAAGTTACTGATCTTTATTAAGAAACCAAGAA 55064
 61 AAAACCAATAGACAGTATTATTATTCATTCGATCGTTTGTCTTATGTTGCTTTC 120
 55063 AAAAATCTCGTTATATGACAAATCGTATTACTCTTATCCAGTTATATTATTAAGTCGC 55004
 121 ATGTGG---CGGAAAGAGCGAAACCGCAAGTCAAAACGATTTGTGACGAGTCTC 177
 55003 ACATGGGTACACGTACGAAAGCGTAGACCAAGTAACATGCTGAGAAAGAAAAATTG 54944

178 TCAGCCTATTGTAGGCGATATTATGATGGAGCGCTATGTTGAAAAAGTAACGAGCA 237
 54943 ACATGACGATGCTTGGGACATTTATGATGAGGACGCTACGTAAGAAAGATTTGTAATG 54884
 238 AAAGGGGACAGACGATATTTTCAATATGTTGAACCGATCTTTAGAGCTCGATTTATGTA 297
 54883 TACGTTACAGATTATGTTTTCGTCATGTTCCGCAATTTTAAAACTCAGATTACGTA 54824
 298 GCAGAAACTTTGAAAAACCC-----GGTAACCTATCAAAAGAAATTAACCAACAGAT 351
 54823 AGTGGAAATTTGAAACATCTGTTTGTGTAAGATTAAGAAATTAACCAACAGAT 54764
 352 AAAGATTTCAATCTGACAGCAAGTAAGAAATCAGTAAAGTCTTGAAGATTAATTC 411
 54763 AAGATTTATCTACTTAAGTCGCAAAAGAAAGAAACGTTAGGACGTAAGAAAGCCGATTT 54704
 412 ACGTTCTCAACAGCGCCACCAACGCAATGATTAACGCGTTACGGGCAATGAAGAT 471
 54703 ACAGATTAAATTTGGCGAATACCATATGACGATTTATGTTGCTAAGGAACTAAAGAT 54644
 472 ACGTTGAGAAATTTGCGAAGCAAACTTGATATCTGTTGAGCGGATACAGCTTAAGT 531
 54643 ACATTAAGGCTTTTAAAGAGCTGATCTTGACTATGTTGGGCTGTGTAATTTCAAA 54584
 532 GATGCGAAAAAGAAATTTGATGACAGAAAGTCAACGCGGTATGCAACGCTTGGC 591
 54583 GATG---TAAATAATTTGTTGATCAAAATGTAATGTTTGGGTTGCTACTTCTTGA 54527
 592 TTACCGATGTTGTCGCGGAAAGTTTCGCGCTAAAGAAATACCGCGGCTGCTGCC 651
 54526 TTACAGATGATTTGATGACAGAGCTATTCGCAAGAAAGAAACAGGTTGTTAAGT 54467
 652 GCAGATCTGAAATCTTATCTTATGATTTGATTAAGAGCAAAAAAC-----AT 699
 54466 ATGAACCCGATGATTAATCTTAAGCAAAATTAAGCAAGATCTTAAAAAGGTAAT 54407
 700 GGTGACATTTGTTGTGACGATCAACATCGGGGCAAGATATGACATGCAACGAC 759
 54406 GGTGATCTTGTGTTGATTAATAGCACTGGGGGAAATAGATTAATTAACGATCTCT 54347
 760 CGCAGCGCCAGCTTTCAGAGACCATGCTGATGCGGAGCTGACATCATCTGCGCAT 819
 54346 AGACAGAGAGCTTTCAGAAACCAATGTTGATGACAGGGGAGATTAATTTGGGACAC 54287
 820 CATCCGACAGCTTTAGAACGATGAAAGTAAATTAACGAAACCGTCAATTTTTCACGCTC 879
 54286 CATCCGATGATCTTCAATCTTTGATGTTGATTAAGCAAGGATTAATCTTAATAGTTTA 54227
 880 GGCACATTTGCTTTGACCAAGGCTGACAGAAACAGAGACAGTCACTGGTTCAGTAT 939
 54226 GGTAACTTTGTTGTTGACCAAGATGACAAAGAAAGATTAAGTCACTGTCAGATAT 54167
 940 CACTGGAAGAAATGAAGCAGCGCTTTGAAGTGAACCAAGTCAATTCATGATGAAGCG 999
 54166 CATTTACGATTAATGATGCTGCAATCTTGATGTTGTTACCTTAATTAATTCAGAGGGA 54107
 1000 ACACCTGCACTGTGTA 1015
 54106 TCACCAAAACCAAGTTA 54091

RESULT 13
 AE017335/c 94830 bp DNA circular BCT 09-JUL-2004
 LOCUS Bacillus anthracis str. 'Ames Ancestor' plasmid px02, complete
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus anthracis str. 'Ames Ancestor'
 Bacillus anthracis str. 'Ames Ancestor'
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus

```
REFERENCE
AUTHORS      1 (bases 1 to 94830)
               Ravel,J., Raako,D.A., Shumway,M.F., Jiang,L., Cer,R.Z.,
               Federova,N.B., Wilson,M., Stanley,S., Decker,S., Read,T.D.,
               Salzberg,S. and Fraser,C.M.
TITLE        Bacillus anthracis comparative genomes
JOURNAL      2 (bases 1 to 94830)
AUTHORS      Ravel,J., Raako,D.A., Shumway,M.F., Jiang,L., Cer,R.Z.,
               Federova,N.B., Salzberg,S. and Fraser,C.M.
TITLE        Direct Submission
JOURNAL      Submitted (17-MAY-2004) Microbial Genomics, The Institute for
               Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850,
               USA
REFERENCE
AUTHORS      3 (bases 1 to 94830)
               Ravel,J., Raako,D.A., Shumway,M.F., Jiang,L., Cer,R.Z.,
               Federova,N.B., Wilson,M., Stanley,S., Decker,S., Read,T.D.,
               Salzberg,S. and Fraser,C.M.
TITLE        Direct Submission
JOURNAL      Submitted (09-JUL-2004) Microbial Genomics, The Institute for
               Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850,
               USA
REMARK
COMMENT      Sequence update by submitter
FEATURES
SOURCE       On Jul 9, 2004 this sequence version replaced gi:47552342.
               Location/Qualifiers
               1..94830
                /organism="Bacillus anthracis str. 'Ames Ancestor'"
                /mol_type="genomic DNA"
                /strain="Ames Ancestor; A2084"
                /db_xref="taxon:261594"
                /plasmid="pXO2"
                /note="MLVA types GT62 and A3_b"
                /complement (543..1139)
                /locus_tag="GBAA_px02_0001"
                /complement (543..1139)
                /locus_tag="GBAA_px02_0001"
                /note="Identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein, (px02-03)"
                /protein_id="AAT28931.2"
                /db_xref="GI:47552343"
                /translation="MTYVKLFQFRGNQIHVDIVDPQRPQYEPREADVLRNHAIVLLD
                EKREKGGODDVLSLKKEVKDLKOSVNTMKQOMDVLIKLMMELIAQSGSNLELBEQ
                EVKSEIASSVTKSBSGKPPRTVREPIEENVOEHRIOREVXTTSFVKKPLPMITDDV
                QIVAKKGRHVQVPRRLCTGYQELMEIDIPHRKKELE"
                /complement (1151..1438)
                /locus_tag="GBAA_px02_0002"
                /complement (1151..1438)
                /locus_tag="GBAA_px02_0002"
                /note="Identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein, (px02-04)"
                /protein_id="AAT28932.2"
                /db_xref="GI:47552344"
                /translation="MEIHVRVDPYHLKEIDRKCKEIGKLGKRYVREYINMFEQI
                FNQYSRNKEDKFBDAVNSVITLDRQSDKLEIYIDATHELVAAVNIKLENE"
                /complement (1652..1768)
                /locus_tag="GBAA_px02_0003"
                /complement (1652..1768)
                /locus_tag="GBAA_px02_0003"
                /note="Identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAT28933.2"
                /db_xref="GI:47552345"
                /translation="MKRKKRYKXHFRIQSAKHRIQAVYEVLKAYKAAHTI"
                /complement (1746..2594)
                /locus_tag="GBAA_px02_0004"
                /complement (1746..2594)
                /locus_tag="GBAA_px02_0004"

gene
CDS
        /note="Identified by Glimmer2; putative"
        /codon_start=1
        /product="conserved hypothetical protein, (px02-05)"
        /protein_id="AAT28934.2"
        /db_xref="GI:47552346"
        /translation="MSIKFMWKEGRKPATKRAVIFLTGSPYTFEFLCISPEBGC
        ATYFEEEMKIGEVQSLSTVIAVAKKEXPNDQTLRIDYELRADNDQSILSNKRYV
        ENKIKOKDNVNTKIVYRASDNVIVISENVPEEFGVSSVVRKPEIYHPELDNDVDL
        KEREMKRVVLENEKTLINRELKKSCKDYERSEYLAFSQALRKETKEKDEKSSAMQDL
        KIKNEQGLTKMEYQTESEKVKITNTINSNSTINNHKEIDVLKEDIKMEKKIKQLLD
        EKKKTI"
        /complement (2591..2866)
        /locus_tag="GBAA_px02_0005"
        /complement (2591..2866)
        /locus_tag="GBAA_px02_0005"
        /note="Identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein, (px02-06)"
        /protein_id="AAT28935.2"
        /db_xref="GI:47552347"
        /translation="MLSEIKPNNDYAKESKYLILSRKKGKVRKQFIPIPTWFDY
        NPEGEVWLIVREYQPSVNGEKXYTNCKLENHQAQSVVNVKGRMK"
        /complement (2955..3563)
        /locus_tag="GBAA_px02_0006"
        /complement (2955..3563)
        /locus_tag="GBAA_px02_0006"
        /note="Identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="conserved hypothetical protein, (px02-07)"
        /protein_id="AAT28936.2"
        /db_xref="GI:47552348"
        /translation="MNNLSKILSGAIVLSIFNGCIAVYNTTPEPDKKKRYRVTSBK
        EKQIGDLKVLBQDKKINSGAGENKKEKNGSTLIDLNKTRBANQVHAHYLVSVON
        KGERRNMLKITDKKQVVDIVAPNTDDIGDNFQSHVAKAIIYNSBDSVSKCTALLD
        IEYTIQELNKQTTINSVVKITILEKQEEIKVEYENYPVKR"
        /complement (3589..4731)
        /locus_tag="GBAA_px02_0007"
        /complement (3589..4731)
        /locus_tag="GBAA_px02_0007"
        /note="Identified by match to Pfam protein family HMM
        PF00877"
        /codon_start=1
        /transl_table=11
        /product="NLP/P60 family protein, (px02-08)"
        /protein_id="AAT28937.2"
        /db_xref="GI:47552349"
        /translation="MAATQETALIDKRYKVRKIWIYVLLGSGNGVIAAATLLIIVS
        MAIRGGSGTGPNGIGSTGATVKNLIPVWRMQAVQECAGAGVELVYVLAIM
        VESNGISEKPLDINQSESGQMANNTISNKDSTVYVWMLTKGAFDAKMLGINDLLA
        IVQTYNFGKRVVHMLAANNKTHSIQTDVYSLTVAVAGGRNGTTGYSQPAVAVYN
        GGYYIYAGNMFVAKWQKLSIFDGAQGTSGQIPGSETRKVMWDEVLKNGNYYVWG
        KSSSGQFDCSGLYVAVYKTAGITIPISATQYDFTEVVPKAPQDVLVEFRGTYGG
        PNHVSHVGIYDANTIMYDSNGSGWGHQFSSVQOHYAGIRVPR"
        /complement (4733..6667)
        /locus_tag="GBAA_px02_0008"
        /complement (4733..6667)
        /locus_tag="GBAA_px02_0008"
        /note="Identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="conserved hypothetical protein, (px02-09)"
        /protein_id="AAT28938.2"
        /db_xref="GI:47552350"
        /translation="MVLIGKKEQKRYTEBGVLDPISRVPOGIGIKENKXYIAGDCY
        VACHVYSLAEDIPMLVLTALMNNKQRTISSVDVITAKKEVVDINSITELKQRTKS
        ERBSTDDDAHWEHLQNLTPARSTTQGGATVYKRIYIDPLDLEKRIKIDKTES
        IAGCNVYQAVYVPEKQEKEMOTLPASVDOQIEYGVSKGYPLPSNIGIPGIPHHQDLK
        DPGIYIGQYSTGSAFTLDPEFSTGRTTSRSGFTFGMGAKSKTLMLQLEBGLVAKOC
        FIRGFDAKRDYVTVQOQGGKIIDLSQVQSEEVETMNPLEVFAKRIWTVGAVDEK
        GSFYQHSKYVTNMIRPLNTEFDDTVIOEFRKHLVAFYIEWGLVQKGSDBRYKVTGYP
```


CDS
/plasmid="PX02"
/complement(19..477)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-01"
/protein_id="AAFI3606.1"
/db_xref="GI:6470152"
/translation="MRGKGDDVSLKKEVKD.KOSVNTMCOMDVLKIMMELIAG
SGNSLSEKLEQEVKSEIASSVTKSEKPRPTKREPIENVEHORIQEVKTTSE
VKRLPMITDDVQIVMKKQPHQVPRLTGQIEWEQIPERRKELE"
/complement(482..1018)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-02"
/protein_id="AAFI3607.1"
/db_xref="GI:6470153"
/translation="MGKSPSIIITSOYTI PNKSFPTTVVDYCMRQKALLKDGHTLTNT
ELTSLQLQNAIDSPDMRGS PHNQKKVSSEKAEAKVLLTSPNSHTTDDKXY
SYMGRKALSKKKKTKTEEEVSLNHIIDLSPBEKKEGKGIRAEITLSCVPSK
DKNVITSTDLNEIKDTIR"
/complement(1037..1633)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-03"
/protein_id="AAFI3608.1"
/db_xref="GI:6470154"
/translation="MTYVKLFOPRGNQIHVDIVDFQRPPEYREADVLRHATVLD
EKMEGKGDDVSLKKEVKD.KOSVNTMCOMDVLKIMMELIAGSGNSLSEKLEQ
EVKSEIASSVTKSEKPRPTKREPIENVEHORIQEVKTTSEVKRLPMITDDV
QIVMKKQPHQVPRLTGQIEWEQIPERRKELE"
/complement(1645..1932)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-04"
/protein_id="AAFI3609.1"
/db_xref="GI:6470155"
/translation="MEIHVRNVDPYHLKEIDRKCKEIGKLGRRYYMEYINMMPEQH
FNQGRSRKEDKPEBAVSNVITLDRQDKQYVIDAHELVAAIMIKLENE"
/complement(2240..3088)
/note="similar to Dictyostelium discoideum
glutamine-asparagine rich protein encoded by GenBank
Accession Number U07817"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-05"
/protein_id="AAFI3610.1"
/db_xref="GI:6470156"
/translation="MSIKFMNKGGRKRPANTKRAYLTLSFVTMFPVLCTSPVFSG
ATYFEBMKIGFYOSISTVKIAAKKREYNDNQLRTDYELRADNSQISNNKYKY
ENKTIKQDNVAKTVRASNDYIVVISENPEEFGVSSVAKPEYIHPBLQNDVLDL
KERSMKYVLNENKLIINRELKKSQDFERELAFVSOGLRKEI EKKMEDSSAMKOL
KIKNEQLTKEMEYQTEGEEKVTKNTINSNESTINNHQKEIDLKEDIKKEKKIQLLD
EKKKTI"
/complement(3085..3360)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-06"
/protein_id="AAFI3611.1"
/db_xref="GI:6470157"
/translation="MLSELKPNHDYAKEGYLLISLRKKKGVRKKKFTETPTWFDY
NFGKEVWLVIVREYQPSVNGKEKYNTCLENIHAQVSVVNGERMK"
/complement(3449..4057)
CDS

CDS
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-07"
/protein_id="AAFI3612.1"
/db_xref="GI:6470158"
/translation="MNNLSKILSGALVLSFNGIWAANYNTTKSEFDKKKKRYTVSEK
EKOIGDLKVLTEQDKKINEGAGENKGNSTLDLONKYREVAOCPVHAYLDYSQV
KGRERNNLKITTRKVDIVAPNTDLDGDNPFKSHVKAIIYINSEGDVSKCTALLD
IETIIELEKQTTINSVKTITLEKQEEELKIVENYPPYKR"
/complement(4083..5225)
/note="similar to Enterococcus faecalis transposon TN916
ORF14 encoded by GenBank Accession Number AB014488"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-08"
/protein_id="AAFI3613.1"
/db_xref="GI:6470159"
/translation="MAATQETALIDKYKVRKIKIVYLLGSGTVVIAAITLLLIYS
MAIFGSGSTGNGGISGTATYKNIPEYVRAQOAVBOCAQGVLELVPYLAITM
VESNGISEKIPDINQSESGMAMNTISNPKDSIYGYVHLKGFADAKLGIINDLLA
IVQYVNGRNVVHMLAANKTHSIQIADYISLVVAPAGNRKGTITGISQPVAVAN
GGYRILINGNFFYAENVKQYLSFDGAGTSGQIPGSEETKVMWDEVLKYNQNPYWG
GKSSQGFDCSGLYVWAKYTAGITIIPISSAATQYDFVTEVDPKADQPDGLVFFGTGYS
PNHSHVGIIYIDANTWYDNSNGSGVGHQFTSSWQGHVAGIRRVPR"
/complement(5227..7158)
/note="similar to Staphylococcus aureus transfer complex
protein trbE encoded by GenBank Accession Number L11988"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-09"
/protein_id="AAFI3614.1"
/db_xref="GI:6470160"
/translation="MKLGKKEQQRRTGYDLDFISRVPOGGIKFNEXYIAQDCYV
ACHLVYSIADIDPLMTLALMNNKDTISSVDVATANKEEVVKDINRITELKRMRTSE
RSTDRDDADHMLQNLTFDPARSITQOEIVKLVKRAYIVDPVLEQLEKIGDKKEI
ACQNYKQVYVFKQKEWMTLPASYSIOIEIVGSGVPLPSKNIGIGIPFHODLD
PRGIYLGQITSGAFILDPFSTGTRISBSGFIIRKMGAKSTLLKLESGIAKDDF
IRGFDKARDVYTVVQOQSGKIIDISGVQDESEVETGINLEVPATKINTGAVDEG
SFIOHISKVTNMRFLNTEEDDTVIOGFRHLVAFYIEMGLVPQKSDRKYKVGYP
NHYPILEDPYKYLNLKLEAGATPQRRHDEALIKLQVEADITVYGDMFNHGLTKNPE
NEQIVPFIDIGISIKSVGSCNCOFTLTLIWSHALNGRQMKIRREKSLSIDVXY
FMVPLDSCHAVINSQNEFANRYVLEFERMRKFPAGIFPATQSPQEMLPBGSTVDVS
TMKAIFFELTOYKIFLANNDSVLETKSVLSESLTSEFRLLPELRGBAIVQVSTET
YVWFPDEKQLEKFKGQ"
/complement(7178..7846)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-10"
/protein_id="AAFI3615.1"
/db_xref="GI:6470161"
/translation="MKVATAEAKVNRKQIESEKSGKOSFISTIPASPKNTDASIV
DIAYKRLBKQTLGVDRNNLOVLYLTKTDLISNODDLKPPMNOULTSCRYVHP
FKILSLYSTETTEQOYTKMKALRPGGRSOSSEKEKHINVQIRSLAIEINRLTL
WVENNELALFFIYVYKNETELIKVQDKMRIGROFNLQNKAKEVEKILFQLQNM
NSEW"
/complement(7910..8224)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-11"
/protein_id="AAFI3616.1"
/db_xref="GI:6470162"
/translation="MEMMRNPKTKTKQIKALFPYIITGAILALMLVLAASYMPKVPYV
GFRIMFYVLFGITGLFLCIKPHNSLPTNRIFFILDLKQDNKYHPIEIVNTISSEK
RK"
/complement(8227..8430)
CDS

DUEAALICECAVOPDYQIIFONKMIQANVIVNVLEHMDVWPTLDEAFAFA
 TIPYNGHLVITSEYLDYFKEVEBERNTKIVADNSRSEFEFLKFDVMPDVAASIA
 LAVEALGIDETEAFRGLNANHPDGMRIIRPADOSRFEFVNPANPDSSLRIM
 ERVDFPGSNLAPVIMNCRPDRDRTEQFARDLPYIKATIVAIGETTAPITSARE
 KGDIPTOBYWNLBEGMSTSEIWSRMKPYIKNIIVGVGNIHGAPELIDMTWEOIGKK
 QAKVI"
 CDS
 1490..1939
 /note="16 Kd encapsulation protein CapC"
 /codon_start=1
 /transl_table=1
 /protein_id="AAA2287.1"
 /db_xref="GI:142632"
 /translation="MFSGDLYIALVLAVTLSTLFTERTGLPAGLVWPGYALVNP
 VEMLVGIVLSITLVIVYGVSRFVLIGRKPATLTGLCNLFEPYCPVMPFEI
 FEPFGIVIVGLIANTIQROGLPLTITGLTLLSGATFANNIYLL"
 1951..3186
 /note="46 Kd encapsulation protein CapA"
 /codon_start=1
 /transl_table=1
 /protein_id="AAA2288.1"
 /db_xref="GI:142633"
 /translation="WRRLTFOBKLLIFIKTKKKNPYVAIVPLVILIAATWY
 RTEAVAPVHRHENEKLTMTWVDIMMGHVAEIVNRVGTVDYFHVSEYLLKSDVSG
 NFEHPVLDEKKNYQKADKNLHLSAKEETVAVEAGETVNLANNHMTDYGATKD
 TIKAFKEADLYVGAGENFKDKNIIVQNVGVAVATLGTFDAFVAGAIATKEQPSGL
 SMNPDLIKOISKADPKKGNADLVVNTTHGEEYDNKSPROELAKAMVDAGDIL
 VGHHPHVLQSPDYKQGIIFYLSIGNFVDDQMTTRKDSALVOYHLRDGTADILVPL
 NIQESPEKPVASALDKNRVYRDLTDSKGLMSKKDKLEIKLNHRIIVKMKREK
 QEHODKQKENVSEVETTT"
 ORIGIN 748 bp upstream of HindIII site.
 Query Match 21.9%; Score 250; DB 1; Length 3244;
 Best Local Similarity 55.6%; Pred. No. 1.9e-52;
 Matches 575; Conservative 0; Mismatches 435; Indels 24; Gaps 4;

Db 2491 GATG--TAAATAATGTGTATCAAAATGTAATGCTGTAGGCTTCTACTCTTGA 2547
 Oy TTTACCATGTGTCGGGAAGGTTTTCGGCTTAAAGAATACGCCGGCGCTGCCCC 651
 Db 2548 TTTTACATGATCTTTTGAAGAGAGCTATTGCAACGAAGAACACAGGTTTGTAGT 2607
 Oy 652 GCAGATCTGTAATCTTCATCCCTATGATTCAGAAACGAAAAA-C-----AT 699
 Db 2608 ATGAACCAAGATGATTAATTAAGCAAAATAGTAAGCAAGATCCTAAAAAGTAAT 2667
 Oy 700 GCTGACATTTGTTGTGTCAGTCACTATGGGGCCAGAGATGACATGATCCAAAGAC 759
 Db 2668 GCTGATCTTTCGTGTAATAGCACTGGGGGGAAGATACATTAATTAACGAGTCT 2727
 Oy 760 CGCCAGCGCAGCTTGCAAGAGCCATGTCATGCGGAGCTGACATCTGTGCGCCAT 819
 Db 2728 AGACAGAGAGCTTGAAGAAAGCAATGTTGATGACAGGCGCATATTTGTGGACAC 2787
 Oy 820 CATCCGACGCTCTTGAACCGATTTGAAGTATTAACGAAACCGTCATTTTCTACAGCCTC 879
 Db 2788 CATCCGATGTAATCTTCAATCTTTGATGTGTATAGCAAGGATTAATCTTATAGTTTA 2847
 Oy 880 GGCACCTTTGCTTTGACCAAGCTGAGACGAAACAAGACAGTGCATGCTCAATAT 939
 Db 2848 GGTAACTTTGTGTTGACCAAGATGACAAAGATGACAAAGATGATGCACTTGCAATAT 2907
 Oy 940 CACCTGAAGAAATGGAACAGCGCGCTTGAAGTGAACCGATCGATATCCATGAACG 999
 Db 2908 CATTAGCTGATTAATGTAATGCAATCTTGTATGTTTACTTTAATATTCAGAGGGA 2967
 Oy 1000 ACACCTGCACCTGT 1013
 Db 2968 TCACCAAAACCAAT 2981

Search completed: February 27, 2006, 11:11:11
 Job time : 6570.47 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 6434.26 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-3

Sequence: 1 atgaaaaaacgacgagctt.....gtgacaactaaatctaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 239354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_hic:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_esc7:*
9: gb_gsa1:*
10: gb_gsa2:*
11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
C 1	45	3.9	1101 10	CNS0039G
2	43	3.8	471 7	CN774916
3	43	3.8	514 7	CN776311
4	43	3.8	790 10	CZ545842
5	43	3.8	1201 10	CNS0164Y
6	42.8	3.8	894 10	CLE510201
7	42.4	3.7	1101 10	CNS0181N
8	42.2	3.7	1020 10	AL564009
9	41.8	3.7	621 3	BI418173
10	41.6	3.6	342 1	AM740409
11	41.4	3.6	587 6	CA301908
12	41.4	3.6	694 7	CV151557
13	41.4	3.6	694 7	CV465225
14	41.4	3.6	534 5	B0844290
15	41.2	3.6	611 5	B0864323
16	41.2	3.6	643 5	B0854399
17	41.2	3.6	890 5	BX441261
18	41.2	3.6	1101 10	CNS0068P
19	41	3.6	920 10	CZ507711
20	41	3.6	930 10	CZ522718
21	40.6	3.6	738 10	CE35032
22	40.6	3.6	801 11	CR184582

C 23	40.4	3.5	724	10	CNS00780	AL066553 Drosophila
C 24	40.2	3.5	376	10	CE759802	CE759802 tigr-gss-
C 25	40.2	3.5	660	5	B0000205	B0000205 QG324D08.
C 26	40.2	3.5	669	5	B0857516	B0857516 QG37L17.Y
C 27	40.2	3.5	694	5	B0863273	B0863273 QG323H02.
C 28	40.2	3.5	703	5	B0861433	B0861433 QG318J09.
C 29	40.2	3.5	726	5	BQ995701	BQ995701 QG310L18.
C 30	40.2	3.5	744	5	BQ864356	BQ864356 QG26H09.
C 31	40	3.5	342	2	BG360474	BG360474 BR110646
C 32	40	3.5	360	1	AM740410	AM740410 BR110553
C 33	40	3.5	572	11	CR298214	CR298214 mrel-15L1
C 34	39.8	3.5	1191	3	BM471276	BM471276 AGENCOURT
C 35	39.6	3.5	270	5	BQ465530	BQ465530 HU03020R
C 36	39.6	3.5	542	5	BQ851926	BQ851926 QG316K16.
C 37	39.6	3.5	549	5	BQ851717	BQ851717 QG316B19.
C 38	39.6	3.5	562	5	BQ865840	BQ865840 QG36A13.Y
C 39	39.6	3.5	570	5	BQ992329	BQ992329 QG25F10.
C 40	39.6	3.5	581	5	BQ853621	BQ853621 QG321A03.
C 41	39.6	3.5	582	5	BQ852156	BQ852156 QG317E12.
C 42	39.6	3.5	591	5	BQ857564	BQ857564 QG37N18.Y
C 43	39.6	3.5	593	3	BI340328	BI340328 36558 MA
C 44	39.6	3.5	593	5	BQ869460	BQ869460 QG56D20.Y
C 45	39.6	3.5	599	5	BQ995224	BQ995224 QG36G14.Y

ALIGNMENTS

RESULT 1
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TERT3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921 GI:4941778
VERSION AL063921
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 1101)
AUTHORS Drosophila melanogaster
TITLE BUKARYOTA, Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

COMMENT Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segrete@genoscope.cns.fr - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosawa and Aaron Mammmer in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bdgpc.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_11b="RPCI-98"
/note="end : TERT3"

ORIGIN

Query Match 3.9%; Score 45; DB 10; Length 1101;
Best Local Similarity 14.9%; Pred. No. 0.55;
Matches 62; Conservative 202; Mismatches 148; Indels 5; Gaps 1;

QY 1 ATGAAAAAGAACTGAGCTTTTCATGAAAAAGCTGTAAGCTGACAAAACAGCAAAAAAG 60
DB 1041 WWWWWATMTDWTMDKMMWWATTAKTDTAMWRTAMRADWAGRDGAGKRDADATDAG 982
QY 61 AAAACCAATAGACAGATTTATTTGCGATTCGATGCTTTTGTCTATGTGCTTTC 120
DB 991 AGRRGGGKRKDK 922
QY 121 ATGTGGCGGGGAAAAAGCGGAAAGCGGAAAGCTATTTCTGACAGCTACTCTCA 180
DB 921 ADDDDGADDDKDDDKGKXADDDTDTGDDDKDKDKDKDKDKDKDKDKDKDKDKDKDK 862
QY 181 GCCTCATTTTGGGATATTTATGATGGAGCGCTATGTTGAAAAAGTAACGAGCAAAA 240
DB 861 GWADADMTWDAADWMAADWRMDAMWMDAMWMDAMWMDAMWMDAMWMDAMWMDAMW 802
QY 241 GGGGCGAG----ACGATATTTTCAATATGTTGAACCGATCTTTAGACCTCGATATG 295
DB 801 DKRADDKRDADDDDAATWTTTTRDTTDDKMKMTDTWTRMAADRTWDRDDDDDD 742
QY 296 TAGCAGAACTTTGAAAAACCGGTAACCTATCAAAAGATTAACAAGCAGATAAG 355
DB 741 RAATGAGKRRRTWKRRKRRRTRMDADADTARDDRRRGGDGAADKGGKTKRRRR 682
QY 356 AGATTCATCTGCAGACGATAGAGATGAGAACTTTGAAAGATATGAAATTTCA 412
DB 681 DRATWDRDAMWADAAMWTTTDTTDDDKDRRRKRRRRRTTARAAMDWMTWKA 625

RESULT 2
CN774916 471 bp mRNA linear EST 20-MAY-2004
LOCUS cae75d10.y1 Hydra EST Darmstadt I Hydra magnipapillata cDNA 5'
DEFINITION similar to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN. ;, mRNA
sequence.

ACCESSION CN774916 GI:47545550
VERSION CN774916
KEYWORDS EST.
SOURCE Hydra magnipapillata
ORGANISM Hydra magnipapillata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
Hydridae; Hydra.
1 (bases 1 to 471)
Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
Martinez,D., Kibler,D., Hampson,S., Clifton,S., Page,D., Marra,M.,
Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagarishvili,R.,
Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
Washu Hydra EST Project
Unpublished (2002)
Contact: Hans Bode
Washu Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Corina Guder / GATC Konstanz, Germany
Library materials provided by Thomas Holstein / Molecular Cell
Biology, TUD, Darmstadt DNA sequencing by: Washington University
Genome Sequencing Center For information on obtaining a clone
please contact: Hans Bode (hnbode@uci.edu)
Seq primer: -40UP.

FEATURES
source location/Qualifiers
1..471
/organism="Hydra magnipapillata"
/mol_type="mRNA"
/strain="sf-1 mutant of Hydra magnipapillata"
/db_xref="taxon:6085"

ORIGIN
Query Match 3.8%; Score 43; DB 7; Length 471;
Best Local Similarity 49.8%; Pred. No. 1.6;
Matches 109; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 691 AAAAACAATGCTGACATTTGTTGTCAGTCACACTGGGGCCAAAGTATGACATGAT 750
DB 216 AAAAGTAAGGTGATATTTATTTGTAACATTTTACACGGGGAAACGAAATCTTGTCTATT 275
QY 751 CCAACGACCGCGCCCGCCAGCTTGCAGAGCCATGCTGATGGGAGGATGATCATC 810
DB 276 CCAAAAGAAACCAAAAGAAATTTGGCAATTTATCTTGTCAATTTGGGTGTAATTTAAATC 335
QY 811 GTGGCCATCATCCGACGCTTTGAAACGATTAAGTATATAAGGAAACGTCATTTTC 870
DB 336 ATGTGACCAACCCACATGTAATGCAAGACATGAATGCGTGAACATAGCGTTGTCAT 395
QY 871 TACAGCTCGGCAACTTGTCTTTGACCAAGCTGACG 909
DB 396 TATAGTTGGGAATCTGTTTTCATTCACATTTACG 434

RESULT 3
CN776311 514 bp mRNA linear EST 20-MAY-2004
LOCUS cae79ef10.y1 Hydra EST Darmstadt I Hydra magnipapillata cDNA 5'
DEFINITION similar to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN. ;, mRNA
sequence.

ACCESSION CN776311 GI:47546945
VERSION CN776311
KEYWORDS EST.
SOURCE Hydra magnipapillata
ORGANISM Hydra magnipapillata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
Hydridae; Hydra.
1 (bases 1 to 514)
Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
Martinez,D., Kibler,D., Hampson,S., Clifton,S., Page,D., Marra,M.,
Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagarishvili,R.,
Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
Washu Hydra EST Project
Unpublished (2002)
Other ESTs: cae79ef10.x1
Contact: Hans Bode
Washu Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Corina Guder / GATC Konstanz, Germany
Library materials provided by Thomas Holstein / Molecular Cell
Biology, TUD, Darmstadt DNA sequencing by: Washington University
Genome Sequencing Center For information on obtaining a clone
please contact: Hans Bode (hnbode@uci.edu)
Seq primer: -40UP
High quality sequence stop: 514.

FEATURES
source location/Qualifiers
1..514
/organism="Hydra magnipapillata"
/mol_type="mRNA"
/strain="sf-1 mutant of Hydra magnipapillata"
/db_xref="taxon:6085"
/lab_host="Transformax EC100 (Epicentre), T1 Phage
resistant electrocompetent cells"
/clone_lib="Hydra EST Darmstadt I"
/note="Vector: pBluescript II SK (+); Site_1: NotI;
Site_2: EcoRI"

ORIGIN

/note="Vector: pbluescript II SK (+); Site_1: NotI;
Site_2: EcoRI"

Query Match 3.8%; Score 43; DB 7; Length 514;
Best Local Similarity 49.8%; Pred. No. 1.6;
Matches 109; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 691 AAAAAACATGCTGACATTTGTTGTGACGACACCTGGGCGCAAGATGACATGAT 750
DB 227 AAAAGTAAGTGAATATTTGTAACATTTTAACTGGGGAACAGAAATCTTTCAT 286
QY 751 CCAAGCAGCCGACGCGCAGCTTGCAAGAGCCATCTGATCGGAGACTGACATCAT 810
DB 287 CCAAGAGAAACCAAGAAATTTGCAATTTTATCTTATGCAATTTGGGTAAATTAATC 346
QY 811 GTGGGCAATCATCGGACGCTTTAGAACCGATTGAAGTATATACGGAACCGTCAATTTTC 870
DB 347 ATTTGAGCCACCCACATATATGACAGACATGAATGCTGAAACATACGCTTGTTCAT 406
QY 871 TACAGCTCGGCACTTTGCTTTTGAACCAAGCTGAGC 909
DB 407 TATAGTTGGGAATCTTTTTCATCCACATTTTCACG 445

RESULT 4
CZ545842 790 bp DNA linear GSS 13-MAY-2005
LOCUS CZ545842
DEFINITION SRAA-aad64c12.g1 Strongyloides ratii whole genome shotgun library
(SRAAGS 004) Strongyloides ratii genomic, genomic survey sequence.
ACCESSION CZ545842
VERSION CZ545842.1 GI:64679122
KEYWORDS GSS.
SOURCE Strongyloides ratii
ORGANISM Strongyloides ratii
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimidae; Strongyloidiidae; Strongyloides.
1 (bases 1 to 790)
REFERENCE Miteva,M., McCarter,J.P., Thompson,F., Viney,M., Page,D.,
Ritter,E., Martin,J., Wylie,T., Dante,M., Waterston,R.H.,
Clifton,S.W. and Wilson,R.
TITLE Genome Survey sequences from the rat parasitic nematode
Strongyloides ratii
JOURNAL Unpublished (2005)
COMMENT Washington University in St. Louis
Contact: Miteva M
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA was provided by Fiona Thompson
(F.Thompson@bristol.ac.uk) and Mark Viney
(Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol,
UK.

FEATURES

source

Class: shotgun.

Location/Qualifiers

1..790

/organism="Strongyloides ratii"

/mol_type="genomic DNA"

/strain="isofemale line ED321 heterogonic"

/db_xref="taxon:34506"

/dev_stage="infective larval stage (L3)"

/lab_host="GSI0"

/clone_1lb="Strongyloides ratii whole genome shotgun
library (SRAAGS 004)"/note="Vector: POTW13; Site_1: BstXI; Site_2: BstXI;
Strongyloides ratii genomic DNA was randomly sheared,
end-repaired and size fractionated to enrich for 2-4 kb
fragments. Genomic DNA was provided by Fiona Thompson
(F.Thompson@bristol.ac.uk) and Mark Viney
(Mark.Viney@bristol.ac.uk) at the University of Bristol,
Bristol, UK. Sequencing by Washington University Genome

ORIGIN

Sequencing Center, St. Louis, MO."

Query Match 3.8%; Score 43; DB 10; Length 790;
Best Local Similarity 62.6%; Pred. No. 1.8;
Matches 67; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 815 GCCATCATCGCAGCGCTTTAGAACCGATTGAAGTATATACGGAACCGTCATTTCTACA 874
DB 7 GGCACATCCGCACTCATTTAGCCCTTGGAATCTACAAAGACGCTCATTTCTACA 66
QY 875 GCGTCGCACTTTGCTTTTGAACCAAGCTGAGCAGAAAGAGAC 921
DB 67 GCGTCGCACTTTGCTTTGCGCTCGGCAATAGCAAGCGAAGGC 113

RESULT 5
CNS0164Y/c 1201 bp DNA linear GSS 26-JUL-1999
LOCUS CNS0164Y/c
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15M04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106300
VERSION AL106300.1 GI:5621234
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1201)
REFERENCE GenomeScope.
Direct Submision
Submitted (23-JUL-1999) GenomeScope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genomecope.cns.fr)
- Web : www.genomecope.cns.fr
determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.dbi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CBRH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES

source

Location/Qualifiers

1..1201

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone_1lb="BACN15M04"

/clone_1lb="DrosBAC"

/plasmid="pBelobAC11"

/note="end : 17"

ORIGIN

Query Match 3.8%; Score 43; DB 10; Length 1201;
Best Local Similarity 35.6%; Pred. No. 2;
Matches 88; Conservative 46; Mismatches 113; Indels 0; Gaps 0;

QY 214 TATGTTGAAGTAAGTACGAGCAAAAGGCGACAGATTTTCAATATGTTGACCG 273
DB 822 TTWATCATTTKATTAATAAAAAAAMAAAAAATAATTTTATTTTTHHWY 763
QY 274 ATCTTAGAGCTCGGATATGATAGCAGAACTTTGAACCCGTAACATCAAAAG 333
DB 762 CAGATTAATTTTDBDAWATTAAGCVRBARRRDRDDDDNDNDWAWATATYCHWH 703
QY 334 AATTATTAACAGCAGATTAAGATTCATCTGCAGACAAATTAAGAACTAGTAAAGTC 393
DB 702 AATATATAGCTCGGCAACAAWAAAGWCTTCTTGCAAAWMAAATAAATTTGKTAW 643
QY 394 TTGAAGATATGAATTTCAAGGTTCTCAACAGGCCAACACGCAATGATTCAGGC 453

	PUBMED	642	ATAGACATTATTAACCTGATATGCACADAAVBTGTCCCMATCAGSANAACATMTAKGT	583
Oy		454	GTTTCAGG	460
Db		582	WTTTATGG	576
RESULT 6				
CLS10201/c				
CLS10201/c				
DEFINITION				
SAIL_825_G02.v1 SAIL Collection Arabidopsis thaliana genomic clone				
SAIL_825_G02.v1, genomic survey sequence.				
CLS10201				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Arabidopsis thaliana (thale cress)				
Arabidopsis thaliana				
Eumariota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicotyledons;				
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi				
1 (bases 1 to 894)				
Seesions,A., Burke,B., Preating,G., Aux,G., McElver,J., Patton,D.,				
Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,				
Bullis,D., Shell,J., Miguel,T., Hutchison,D., Kimmerly,B.,				
Mitxel,T., Katsagiri,F., Glazebrook,J., Law,M. and Goff,S.A.				
A high-throughput Arabidopsis reverse genetics system				
Plant Cell 14 (12), 2985-2994 (2002)				
12468722				
Contact: Seesions A				
Applied Trait Genetics				
Syngenta Biotechnology Inc.				
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA				
Email: allen.seesions@syngenta.com				
ABRC Stock Number CS836839; T-DNA left border flanking sequences of				
Syngenta Arabidopsis Insertion Library (SAIL) lines are available				
through the Arabidopsis Biological Resource Center (ABRC).				
Sequences represent a pool of amplified genomic regions and not				
single contiguous sequences.				
Class: TDNA tagged.				
Location/Qualifiers				
1..894				
/organism="Arabidopsis thaliana"				
/mol_type="genomic DNA"				
/ecotype="Columbia"				
/db_xref="taxon:3702"				
/clone="SAIL_825_G02.v1"				
/clone.lib="SAIL Collection"				
/note="T-DNA left border sequences were isolated using a				
modified TAII-PCR strategy"				
ORIGIN				
Query Match				
Best Local Similarity				
Matches				
80; Conservative				
0; Mismatches				
62; Indels				
0; Gaps				
0,				
Oy				
258	TCAATATGTTAACCGACTCTTAGAGCCTCGGATTATGTAGCAGAACTTGAAAACC	317		
Db				
650	TAAATATTTTGAACCAAGACTTAAACAAGAAGCTATGTTATTTGAAAAAATTGGAACCCCTT	591		
Oy				
318	GCTAACCTATCAAAGAATTATTAACAAGCATTAAGAAGTTCTTCGCAGACGAATAA	377		
Db				
590	GAGAAACATCAAGTATATGCAACCAAAAAGTAAAAAATGATATTGAAAGCCGAA	531		
Oy				
378	GGAATCAGTGAAGTCTTGAAG	399		
Db				
530	GGAAGTGAACAAAGTGAAGAG	509		
RESULT 7				
CNS0181N				
CNS0181N				
DEFINITION				
Drosophila melanogaster genome survey sequence Sp6 end of BAC				
BACN37P10 of DrosBAC library from Drosophila melanogaster (fruit				

ACCESSION	AL108773
VERSION	GI:5629077
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)
REFERENCE	Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (BAC BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
FEATURES	Location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACN37P10" /clone_1lb="DrosBAC" /plasmid="pBelobAC11" /note="end : Sp6"
ORIGIN	Query Match 3.7%; Score 42.4; DB 10; Length 1101; Best Local Similarity 22.6%; Pred. No. 2.9; Matches 83; Conservative 113; Mismatches 171; Indels 0; Gaps 0;
Oy	86 CCATTCCGATCGTTTGTCTTATGTTCCGTCATGTGGCGGAAAGCGAACC 145
Db	579 CCCCCCMANVTTTTTTTTTTTTTTTTTTTTTTTTGTGSSGKGRYBKTTTBKB 638
Oy	146 CGAAGCTCAAAAGCATTTGCAGCAGCTACTCAGCTCATTTGTAGGCGATTTATGA 205
Db	639 GTGTTTTTGGSSTTTTGTGAGGAKTKTKKGKGGTTTTTTTTTTTMMMAAADAR 698
Oy	206 TGGGACGCTATGTTGAAAAATAACGAGCAAAAAGGGCAGACATATTTCAATATG 265
Db	699 GRGVVMGVSVAHKQCKTWMDRTTKTSABRABRAMGRSARMAAARTTHMMAAAAA 758
Oy	266 TTGAACGATCTTAGAGCTCGGATTAATGACAGAACTTTGAAAAACCCGTAACCT 325
Db	759 RAAMHARRAMWTHHRVAPARAGRRGRRGGDBRARBRHRRRRARAANAARRGRRRRARG 818
Oy	326 ATCAAAGAATTATAACAAGCAATATAAGATTTCAATCGACAGCAATAAGAAATCAG 385
Db	819 RRMVVRMRAANNAARAGRAPRAGRAPRARPAPRARAARGRRRRGRRGRRRRAR 878
Oy	386 TGAAGCTTGAAGATGAATTTCAAGCTTCCAACAGCGCCAACACACCAATG 445
Db	879 AARPAARARARARGRRAAARAAAGGAGGAGRAPARPARAGRRGRRRRRRRRRRG 938
Oy	446 ATTACGG 452
Db	939 RRGRAGG 945
RESULT 8	ALS64009/ c . 1020 bp mRNA linear EST 05-APR-2004
LOCUS	ALS64009 Homo sapiens FETTL LIVER Homo sapiens cDNA clone
DEFINITION	CSDNM001YE17 3-PRIME, mRNA sequence.
ACCESSION	ALS64009

source

1. .342
/organism="Biomphalaria glabrata"
/mol_type="mRNA"
/strain="BS-90"
/db_xref="taxon:6526"
/clone="RBGIH55TR"
/sex="hermaphrodite"
/cell_type="Hemocyte"
/lab_host="laboratory host"
/clone_lib="Biomphalaria glabrata (BS-90)-unexposed Lambda Zap Library"
/note="vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Total RNA was isolated from the hemocytes of unexposed Biomphalaria glabrata (BS-90) snails and first strand cDNA synthesized using an oligo-dT primer-linker (XhoI). Second strand synthesis was followed by the ligation of EcoRI adaptors. Following digestion with XhoI, the completed, directional cDNA was cloned into Uni-ZAP XR Phagemid vector by Stratagene."

ORIGIN

Query Match 3.6%; Score 41.6; DB 1; Length 342;
Best Local Similarity 50.0%; Pred. No. 3.7;
Matches 104; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

OY 22 CATGAAAGCTGTCTAAGCTGACAAAGCAAAAGAAACCAATAGACGATTT 81
DB 2 CACGAGCATGATGATGAGAGGTAAAGCTCACGACCTGACCAACATGACATCTTT 61
OY 82 ATTGCCATTCGATGCTTTTCTTATGTTGCTTCATGTGGCGGAAAGCCGAA 141
DB 62 GGTGTATCATGCTCTATTTAGACATCATGATGTTTTCTGTACATCTTATCATTCGCG 121
OY 142 ACGCGAAGGTCAAAAGCTATTTTCACAGACGATCTCAGCTCTATTGTAGGCGATTT 201
DB 122 GACTCGACCTGAGAAACATATTCAGTCTTAACCTCACAATTTCCAAATCAGAAAGAAAA 181
OY 202 ATGATGGGACGCTATGTTGAAAAGTAA 229
DB 182 ATCTAGACAAAGAAACATTTGAAGAACTAA 209

RESULT 11
CA301908/c 587 bp mRNA linear EST 01-NOV-2002
LOCUS caa10d03.x1 Hydra cDNA library Hydra magnipapillata cDNA 3' similar
DEFINITION to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN.; mRNA
SEQUENCE.
CA301908
ACCESSION CA301908
VERSION CA301908.1 GI:24464967
KEYWORDS EST.
SOURCE Hydra magnipapillata
ORGANISM Hydra magnipapillata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;
Hydridae; Hydra.
REFERENCE 1 (bases 1 to 587)
AUTHORS Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q., Martinez,D., Kibler,D., Hampson,S., Clifton,S., Page,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Danne,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagarashvili,R., Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
TITLE WASHU Hydra EST Project
JOURNAL Unpublished (2002)
COMMENT Contact: Hans Bode
Washu Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library was constructed by H. Bode and B. Blumberg DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Hans Bode (hnbode@wustl.edu)

Seq primer: Primer name ambiguous
High quality sequence stop: 404.
Location/Qualifiers

FEATURES
source

1. .587
/organism="Hydra magnipapillata"
/mol_type="mRNA"
/strain="105"
/db_xref="taxon:6085"
/lab_host="DHSalpa"
/clone_lib="Hydra cDNA library"
/note="vector: pSPORT6; Site 1: NotI; Site 2: SalI; Libraries prepared by Phil Wigge, Plant Molecular and Cellular Biology, Lab (Detlev Weigelt), The Salk Institute for Biological Science, 10010 North Torrey Pines Road, La Jolla, CA 92037."

ORIGIN

Query Match 3.6%; Score 41.4; DB 6; Length 587;
Best Local Similarity 49.3%; Pred. No. 4.8;
Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 691 AAAAATGATGTCATTTGTTGTCAGTCACTGCGGCCAAGTATGACATGAT 750
DB 554 AAAAGTAAAGGTATATTATTGTAACATTTTACACTGCGGAAACAGATCTTGTATT 495
OY 751 CCAAGCAGCGCCGCGCCAGCTTGCAGAGCCATGTCGATGCGGAGCTGACATCATC 810
DB 494 CCAAAAGAAACAAAGAAATTTGCGAATTTATCTTATGTCATTTGGGTGAATTTATATC 435
OY 811 GTGCGCATCATTCGCGACGCTTTAGAAACGATTAAGTATATAAGGAAACCGCATTTTC 870
DB 434 ATTGATGACCAACCCACATGTAATGCAAGACATGAGGCGTGAACATACGCTTGTTCAT 375
OY 871 TACAGCTTCGCACTTTGCTTTGACCAAGCGTGCAGC 909
DB 374 TATAGTTGGAAATCTGTTTTCATTCACATTCACG 336

RESULT 12
CV151557/c 694 bp mRNA linear EST 08-SEP-2004
LOCUS ca171e03.x2 Hydra EST UCI 5 ALP Hydra magnipapillata cDNA 3'
DEFINITION similar to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN.; mRNA
SEQUENCE.
CV151557
ACCESSION CV151557.1 GI:51945203
KEYWORDS EST.
SOURCE Hydra magnipapillata
ORGANISM Hydra magnipapillata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;
Hydridae; Hydra.
REFERENCE 1 (bases 1 to 694)
AUTHORS Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q., Martinez,D., Kibler,D., Hampson,S., Clifton,S., Page,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Danne,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagarashvili,R., Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
TITLE WASHU Hydra EST Project
JOURNAL Unpublished (2002)
COMMENT Other ESTs: ca171e03.y2
Contact: Hans Bode
Washu Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library was constructed by Dirk Lindgens, Univ. of Calif., Irvine
Library materials provided by Hans Bode & Dirk Lindgens, Univ. of Calif., Irvine DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Hans Bode (hnbode@wustl.edu)
Seq primer: -40UP from Gibco

High quality sequence stop: 694.
Location/Qualifiers

FEATURES

source

1. 694

/organism="Hydra magnipapillata"

/mol_type="mRNA"

/strain="105"

/db_xref="taxon:6085"

/lab_host="ElectroMAX DH10B cells, Invitrogen"

/clone_1ib="Hydra EST UCI 5 ALP"

/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; a.1st strand cDNA was primed with a Not I primer-adaptor (5' -pGACTGTTCTAGATCGCGCGGCCG(T)15-3')"

b.Double-stranded cDNA was ligated to Sal I adaptor, digested with Not I and cloned into the pSPORT1-vector pre-cut with Not I and Sal I. c.The ligation mix was transformed into DH10B cells. d.The picked clones were grown up in the 384-plate containing the freezing medium (LB-Medium containing 4 % glycerol w/v, and different salts) (antibiotic is carbenicillin 100ug/ml). e.The frequency of vectors containing inserts is 96% as determined by digestion check after picking 20 clones, miniprep and subsequent digestion with Not I and Sal I. f.A low level of 32P was used in the cDNA synthesis procedure. The level measured by holding a Geiger Counter next to a plate was background."

ORIGIN

Query Match 3.6%; Score 41.4; DB 7; Length 694;

Best Local Similarity 49.3%; Pred. No. 5;

Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

691 AAAAAATGCTGACATTTGTTGTCAGTCACTGGGGCCAGATGCAATGAT 750

571 AAAAGTAAGTGAATTTATTTGACATTTTACATGGGGAACAGAAATCTTGCTATT 512

751 CCAAGACCGCCAGCGCCAGCTTGAAGACCATGTCATCGGAGCTGACATC 810

511 CCAGAAAGAAACAAAGAAATTTGGCAATTTATTTAGTCAATTTGGTGAATTTAAATC 452

811 GTGGGCATCATCGGACGCTTAGAAGCATTAATTAAGGAAACCGTCATTTTC 870

451 ATTGTAGCCACCAATGTAATGCAAGACATGAGTGTGAACAATCGCTTGTTCAT 392

871 TACAGCTCGGCAACTTTGTTTGAACCAAGGCTGAGC 909

391 TATAGTTGGGAATCTGTTTTCATTCACATTTTACG 353

RESULT 13 842 bp mRNA linear EST 30-SEP-2004

CV465225/c taj24e01.x1 Hydra EST UCI 5 ALP Hydra magnipapillata cDNA 3'

DEFINITION similar to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN. ; mRNA

sequence.

CV465225 GI:52860155

EST.

Hydra magnipapillata

Hydra magnipapillata

Hydridae; Hydra.

1 (bases 1 to 842)

Bode,H., Blumberg,B., Steele,R., Widge,P., Gee,L., Nguyen,O.,

Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,

Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,

Gibbons,M., Ritzer,E., Bennett,U., Ronko,I., Tsagarashvili,R.,

Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.

Unpublished (2002)

Washu Hydra EST Project

Contact: Hans Bode

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@washington.wustl.edu

Library was constructed by Dirk Lindgens, Univ. of Calif. Irvine

Library materials provided by Hans Bode & Dirk Lindgens, Univ. of

Calif., Irvine DNA sequencing by: Washington University Genome

Sequencing Center For information on obtaining a clone please

contact: Hans Bode (hobode@uci.edu)

Seq primer: -40bp from Gibco

High quality sequence stop: 635.

FEATURES

source

1. 842

/organism="Hydra magnipapillata"

/mol_type="mRNA"

/strain="105"

/db_xref="taxon:6085"

/lab_host="ElectroMAX DH10B cells, Invitrogen"

/clone_1ib="Hydra EST UCI 5 ALP"

/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; a.1st strand cDNA was primed with a Not I primer-adaptor (5' -pGACTGTTCTAGATCGCGCGGCCG(T)15-3')"

b.Double-stranded cDNA was ligated to Sal I adaptor, digested with Not I and cloned into the pSPORT1-vector pre-cut with Not I and Sal I. c.The ligation mix was transformed into DH10B cells. d.The picked clones were grown up in the 384-plate containing the freezing medium (LB-Medium containing 4 % glycerol w/v, and different salts) (antibiotic is carbenicillin 100ug/ml). e.The frequency of vectors containing inserts is 96% as determined by digestion check after picking 20 clones, miniprep and subsequent digestion with Not I and Sal I. f.A low level of 32P was used in the cDNA synthesis procedure. The level measured by holding a Geiger Counter next to a plate was background."

ORIGIN

Query Match 3.6%; Score 41.4; DB 7; Length 842;

Best Local Similarity 49.3%; Pred. No. 5.2;

Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

691 AAAAAATGCTGACATTTGTTGTCAGTCACTGGGGCCAGATGCAATGAT 750

571 AAAAGTAAGTGAATTTATTTGACATTTTACATGGGGAACAGAAATCTTGCTATT 512

751 CCAAGACCGCCAGCGCCAGCTTGAAGACCATGTCATCGGAGCTGACATC 810

511 CCAGAAAGAAACAAAGAAATTTGGCAATTTATTTAGTCAATTTGGTGAATTTAAATC 452

811 GTGGGCATCATCGGACGCTTAGAAGCATTAATTAAGGAAACCGTCATTTTC 870

451 ATTGTAGCCACCAATGTAATGCAAGACATGAGTGTGAACAATCGCTTGTTCAT 392

871 TACAGCTCGGCAACTTTGTTTGAACCAAGGCTGAGC 909

391 TATAGTTGGGAATCTGTTTTCATTCACATTTTACG 353

RESULT 14 534 bp mRNA linear EST 14-AUG-2002

B0844290 CGA13H12.yg ab1 OG_ABDCl lettuce salinas Lactuca sativa cDNA clone

LOCUS CGA13H12, mRNA sequence.

DEFINITION B0844290

ACCESSION B0844290

VERSION B0844290.1 GI:22224840

KEYWORDS

SOURCE

ORGANISM

Lactuca sativa

Lactuca sativa

Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae; Lactuca.

1 (bases 1 to 534)

Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,

TITLE
Liu, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L., and Bradford, K.
Lettuce and Sunflower ESTs from the Composite Genome Project
<http://compgenome.ucdavis.edu/>

JOURNAL
Unpublished (2002)

COMMENT
Contact: Alexander Kozik [R.W.Michelmores]
Department of Vegetable Crops, R.W.Michelmores Lab
University of California at Davis (UCD)
Aamundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmores@vegmail.ucdavis.edu]
belongs to contig OG_CA_Contig6559, see <http://cgpdb.ucdavis.edu/>
for details.
Plate: OG13 row: H column: 12.

FEATURES
source
1. 534
Location/Qualifiers
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="OG13H12"
/lab_host="E.coli"
/clone_lib="OG ABCDI lettuce salinas"
/note="Vector: pBRCDNASF1AB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at <http://cgpdb.ucdavis.edu/>
TAG_SEQ=Not found"

ORIGIN
Query Match 3.6%; Score 41.2; DB 5; Length 534;
Best Local Similarity 50.5%; Pred. No. 5.3; Mismatches 0; Gaps 0;
Matches 100; Conservative 0; Indels 0; Gaps 0;
QY 448 TACGGCGTTGAGGCGATGAAGATACGCTTGAGAAATTTGCGAACAACCTTGATATC 507
DB 309 TCCGCTTCGAGATCTGAAGCATCATGTTGAAGACATCGGATCGCTTTCGATTTTC 250
QY 508 GTTGAGCGGGATACGCTTAAGTATGCGAAAAAATTCTTACCAAGAACTCAAC 567
DB 249 GTCGAAGAGACGACGGGTGAGGGGCGGCGGAAACCGCTTCGTTAACTGACCGCTTC 190
QY 568 GGGGTAAAGATTGCAACGCTTGCTTACCGATGTGTCGGGAAAGTTTCGGGCTTAA 627
DB 189 GGTGTAGCCAACTGAACCGGGCGGTGAACCGATGAGCTTTGAGACGGTGTGCGTTCCAT 130
QY 628 AAGAAATACCGCGGGCGGTG 645
DB 129 GAACCTCACTCATGTGAG 112

RESULT 15
BO864323 611 bp mRNA linear EST 14-AUG-2002
LOCUS BO864323/c
DEFINITION OG26J20.yg.ab1 OG ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION BO864323
VERSION BO864323.1 GI:22249869
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; campanulids; Asterales; Asterales; Cichorioideae;
Cichorioideae; Lactuca.
REFERENCE 1 (bases 1 to 611)

AUTHORS
Kozik, A., Michelmores, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Liu, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L., and Bradford, K.
Lettuce and Sunflower ESTs from the Composite Genome Project
<http://compgenome.ucdavis.edu/>

JOURNAL
Unpublished (2002)

COMMENT
Contact: Alexander Kozik [R.W.Michelmores]
Department of Vegetable Crops, R.W.Michelmores Lab
University of California at Davis (UCD)
Aamundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmores@vegmail.ucdavis.edu]
belongs to contig OG_CA_Contig6559, see <http://cgpdb.ucdavis.edu/>
for details.
Plate: OG26 row: J column: 20.

FEATURES
source
1. 611
Location/Qualifiers
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="OG26J20"
/lab_host="E.coli"
/clone_lib="OG ABCDI lettuce salinas"
/note="Vector: pBRCDNASF1AB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at <http://cgpdb.ucdavis.edu/>
TAG_TISSUE=chemical induction
TAG_LIB=OG ABCDI lettuce salinas
TAG_SEQ=GTGAGCGCGG"

ORIGIN
Query Match 3.6%; Score 41.2; DB 5; Length 611;
Best Local Similarity 50.5%; Pred. No. 5.5; Mismatches 98; Indels 0; Gaps 0;
Matches 100; Conservative 0; Indels 98; Indels 0; Gaps 0;
QY 448 TACGGCGTTGAGGCGATGAAGATACGCTTGAGAAATTTGCGAACAACCTTGATATC 507
DB 332 TCCGCTTCGAGATCTGAAGCATCATGTTGAAGACATCGGATCGCTTTCGATTTTC 273
QY 508 GTTGAGCGGGATACGCTTAAGTATGCGAAAAAATTCTTACCAAGAACTCAAC 567
DB 272 GTGAAAGAGACGACGGGTGAGGGGCGGCGGAAACCGCTTCGTTAACTGACCGCTTC 213
QY 568 GGGGTAAAGATTGCAACGCTTGCTTACCGATGTGTCGGGAAAGTTTCGGGCTTAA 627
DB 212 GGTGTAGCCAACTGAACCGGGCGGTGAACCGATGAGCTTTGAGACGGTGTGCGTTCCAT 153
QY 628 AAGAAATACCGCGGGCGGTG 645
DB 152 GAACCTCACTCATGTGAG 135

Search completed: February 27, 2006, 11:09:29
Job time : 6438.26 secs

or more than two genes of the Bacillus subtilis poly-gamma-glutamate synthetase complex (psbCA) to facilitate microbial surface expression of the target protein. The psbBCA gene complex comprises the psbB, psbA and psbA genes and is normally expressed in the outer membrane of Bacillus subtilis. The vector can be transformed into either Gram-positive or Gram-negative bacteria (e.g., Escherichia coli), and can be used for the surface expression of various proteins of interest such as enzymes, antigens, antibodies, attachment proteins or adsorption proteins. CC proteins recombinantly produced using the vector of the invention can be used as, for example, vaccines or enzymes. The present sequence CC represents the Bacillus subtilis psbBCA complex gene psbA, which is specifically claimed for use in the vector of the invention.

Sequence 1140 BP; 381 A; 235 C; 265 G; 259 T; 0 U; 0 Other;

Query Match 100.0%; Score 1140; DB 10; Length 1140;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTTAAAGCTGACAAAAAGAAAAAG 60
1 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTTAAAGCTGACAAAAAGAAAAAG 60
61 AAAACCAATTAAGCAGTATTTATTCATTCGATCGTTTTCCTTATGTTGCTTTC 120
61 AAAACCAATTAAGCAGTATTTATTCATTCGATCGTTTTCCTTATGTTGCTTTC 120
121 ATGTGGGGGAAAAAGCCGAAAGTCAAAAGTATTCGACAGAGTACTCTCA 180
121 ATGTGGGGGAAAAAGCCGAAAGTCAAAAGTATTCGACAGAGTACTCTCA 180
121 ATGTGGGGGAAAAAGCCGAAAGTCAAAAGTATTCGACAGAGTACTCTCA 180
181 GCCCATTTGTAGAGGATATTAATGATGGAGGCTATGTTGAAGAAAGTAAAGCAAAA 240
181 GCCCATTTGTAGAGGATATTAATGATGGAGGCTATGTTGAAGAAAGTAAAGCAAAA 240
241 GGGGCGACAGATATTTTCAATATGTTGAACGATCTTTAGAGCCTCGGATTTATGACA 300
241 GGGGCGACAGATATTTTCAATATGTTGAACGATCTTTAGAGCCTCGGATTTATGACA 300
301 GGAACCTTTGAAAAACCCGGTAACTATCAAAAGATTTAAACAGCAGATTAAGAGATT 360
301 GGAACCTTTGAAAAACCCGGTAACTATCAAAAGATTTAAACAGCAGATTAAGAGATT 360
361 CATCTGACAGAGAAATTAAGAAATCAAGTAAAGTGAAGATTAATTTACCGGTTCTC 420
361 CATCTGACAGAGAAATTAAGAAATCAAGTAAAGTGAAGATTAATTTACCGGTTCTC 420
421 AACAGCGCAACCAACGACATGATTAACGCGCTTCAAGGCGATAAAGATACGCTTGA 480
421 AACAGCGCAACCAACGACATGATTAACGCGCTTCAAGGCGATAAAGATACGCTTGA 480
481 GAATTTGGAGAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAGTATCGAAA 540
481 GAATTTGGAGAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAGTATCGAAA 540
541 AAGAAAAATTTGCTACAGAAAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 600
541 AAGAAAAATTTGCTACAGAAAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 600
601 GTGTCCGGGAAAAGTTTCCGCGCTAAAAAAGATACCGCGGCGTCTGCGCGACAGATCT 660
601 GTGTCCGGGAAAAGTTTCCGCGCTAAAAAAGATACCGCGGCGTCTGCGCGACAGATCT 660
661 GAATTCCTTCATCCCTATGATTTCAAGAGCGAAAAAATCATGTCGACATTTGTTGCAAG 720
661 GAATTCCTTCATCCCTATGATTTCAAGAGCGAAAAAATCATGTCGACATTTGTTGCAAG 720
721 TCACACTGGGGCCAAAGATGACATGATCCAAACGACCGCGACCGCAAGCTTCAAGA 780
721 TCACACTGGGGCCAAAGATGACATGATCCAAACGACCGCGACCGCAAGCTTCAAGA 780
781 GCCATGCTGATCGGGAGCTGACATCATCGTGGCGCATCATCCGACGCTTTAGAACG 840
781 GCCATGCTGATCGGGAGCTGACATCATCGTGGCGCATCATCCGACGCTTTAGAACG 840
```

```
DB 781 GCCATGCTGATCGGGAGCTGACATCATCGTGGCGCATCATCCGACGCTTTAGAACG 840
QY 841 ATTAAGTATATACGGAACCGTCAATTTTACAGCCTCGGCAACTTGTCTTTGACCA 900
DB 841 ATTAAGTATATACGGAACCGTCAATTTTACAGCCTCGGCAACTTGTCTTTGACCA 900
QY 901 GGCTGACGAGAACAGAGACAGTGCATGTTCTGATACCTGAAAGAAAAATGAGACA 960
DB 901 GGCTGACGAGAACAGAGACAGTGCATGTTCTGATACCTGAAAGAAAAATGAGACA 960
QY 961 GGCGGCTTTGAAGTACACCGATGATATTCATGAGACGACACCTGCACTGTGAAAAA 1020
DB 961 GGCGGCTTTGAAGTACACCGATGATATTCATGAGACGACACCTGCACTGTGAAAAA 1020
1021 GACAGCCTTAAACAGAAACCATATTTCGGGAACGACGAAAGACTTAAATTTGCTTGG 1080
1021 GACAGCCTTAAACAGAAACCATATTTCGGGAACGACGAAAGACTTAAATTTGCTTGG 1080
1081 AAAGTAGAGACGGAAGAACTGACGTTGATATTTGATCATAGTGACAAACTTAAATCTTAA 1140
1081 AAAGTAGAGACGGAAGAACTGACGTTGATATTTGATCATAGTGACAAACTTAAATCTTAA 1140
```

RESULT 2

AD007643 ID AD007643 standard; DNA; 1140 BP.

AC AD007643;

DT 15-JUN-2004 (first entry)

DE B subtilis poly-X-glutamate synthetase complex coding sequence psbA.

KM ds; gene; enzyme; vaccine; cytosolic; psbA; poly-X-glutamate synthetase;

KW human papilloma virus.

OS Bacillus subtilis.

XX WO2004035795-A1.

XX 29-APR-2004.

XX 17-OCT-2003; 2003WO-KR002163.

XX 17-OCT-2002; 2002KR-00063378.

XX (BIOL-) BIOLEADERS CORP.

XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

XX Sung M, Poo H, Lee J, Jung C, Hong S, Kim C, Park S, Pyo H;

XX WPI: 2004-348463/32.

XX New vector containing psb A-C genes encoding poly-gamma-glutamate

XX synthetase complex and an antigen protein gene of human papilloma virus,

XX useful in preparing vaccine for treating or preventing mucosal tumor,

XX e.g. cervical cancer.

XX Disclosure; Page 62-63; 69pp; English.

XX The present invention relates to a vector for preparing a vaccine which

XX contains one or more than two genes, i.e. psb A-C encoding poly-X-

XX glutamate synthetase complex and an antigen protein gene of human

XX papilloma virus. The vector and microbes transformed with it are useful

XX in preparing vaccines for treating or preventing mucosal tumor, e.g.

XX cervical cancer. The present sequence is a Bacillus subtilis poly-X-

XX glutamate synthetase complex coding sequence.

Sequence 1140 BP; 381 A; 235 C; 265 G; 259 T; 0 U; 0 Other;

Query Match 100.0%; Score 1140; DB 12; Length 1140;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAAAAAAGAACTGAGCTTTGATGAAAAGCTCTTAAAGCTGACAAAACAGAAAAAAG 60
DB 1 ATGAAAAAAGAACTGAGCTTTGATGAAAAGCTCTTAAAGCTGACAAAACAGAAAAAAG 60
QY 61 AAAACCAATPAGCAGTATTTATGTCATTCGATCGTGTGCTTATGTTGCTTTC 120
DB 61 AAAACCAATPAGCAGTATTTATGTCATTCGATTCGATCGTGTGCTTATGTTGCTTTC 120
QY 121 ATGTGGCGGGGAAAAGCGGAAAACGCCGAAGTCAAAAGCTATTTGACGACGACTCTCA 180
DB 121 ATGTGGCGGGGAAAAGCGGAAAACGCCGAAGTCAAAAGCTATTTGACGACGACTCTCA 180
QY 181 GCCTCATTTGTAGCGCATATATGATGGAACGCTAATGTTGAAAAGTAAACGAGCAAAA 240
DB 181 GCCTCATTTGTAGCGCATATATGATGGAACGCTAATGTTGAAAAGTAAACGAGCAAAA 240
QY 241 GGGGCGACAGATTTTCAATATGTTGAACGATCTTTAGAGCTCGGATTAATGAGCA 300
DB 241 GGGGCGACAGATTTTCAATATGTTGAACGATCTTTAGAGCTCGGATTAATGAGCA 300
QY 301 GGAACCTTTGAAAACCCGGTAACTTATCAAAAGATTAATTAACAGCAATTAAGAGATT 360
DB 301 GGAACCTTTGAAAACCCGGTAACTTATCAAAAGATTAATTAACAGCAATTAAGAGATT 360
QY 361 CATCTGACAGCAATPAGCAATCAATGTAAGTCTTGAAGATATGATTTACGGTTCTC 420
DB 361 CATCTGACAGCAATPAGCAATCAATGTAAGTCTTGAAGATATGATTTACGGTTCTC 420
QY 421 AACAGCGCCAAACACCAACGATTCAGCGGCTCAGGGCGATGAAGATACGTTGGA 480
DB 421 AACAGCGCCAAACACCAACGATTCAGCGGCTCAGGGCGATGAAGATACGTTGGA 480
QY 481 GAATTTGCGAAGCAAAACCTTGATATCGTTGAGCGGATACAGCTTAAATGATGCGAAA 540
DB 481 GAATTTGCGAAGCAAAACCTTGATATCGTTGAGCGGATACAGCTTAAATGATGCGAAA 540
QY 541 AAGAAAAATTTCTGACAGAAAAGTCAACGGGGTAAACGATTCGTAACGCTTACCGAT 600
DB 541 AAGAAAAATTTCTGACAGAAAAGTCAACGGGGTAAACGATTCGTAACGCTTACCGAT 600
QY 601 GTGTCCGGGAAAAGCTTTCCGGCTTAAAAAATACCGCGGCGTGTGCTCCCGAGATCC 660
DB 601 GTGTCCGGGAAAAGCTTTCCGGCTTAAAAAATACCGCGGCGTGTGCTCCCGAGATCC 660
QY 661 GAATCTCTTCACTCTATGATTTGAGAAAGCGAAAAAACAATCTGACATTTGTTGTGCA 720
DB 661 GAATCTCTTCACTCTATGATTTGAGAAAGCGAAAAAACAATCTGACATTTGTTGTGCA 720
QY 721 TCACACTGGGGCCAGAGATGACAAATGATCCAAACGACCGCCAGCCGCTTGCAAGA 780
DB 721 TCACACTGGGGCCAGAGATGACAAATGATCCAAACGACCGCCAGCCGCTTGCAAGA 780
QY 781 GCCATGCTGATGCGGAGCTGACATATGTTGCGGCAATATCCGACGCTTAAAGACG 840
DB 781 GCCATGCTGATGCGGAGCTGACATATGTTGCGGCAATATCCGACGCTTAAAGACG 840
QY 841 ATTGAATATATTAACGAGACGTCATTTCTACAGCTCTGCGCAACTTTGCTTGAACAA 900
DB 841 ATTGAATATATTAACGAGACGTCATTTCTACAGCTCTGCGCAACTTTGCTTGAACAA 900
QY 901 GGGTGCAGCAAGAACAGAGACAGTGCAGTGTTCAGATGACCTGGAAGAAAATGAGACA 960
DB 901 GGGTGCAGCAAGAACAGAGACAGTGCAGTGTTCAGATGACCTGGAAGAAAATGAGACA 960
QY 961 GGGCGCTTTGAAAGTGAACCGATGATTCATGAGACGACCTGCACTGTGAAAAAA 1020
DB 961 GGGCGCTTTGAAAGTGAACCGATGATTCATGAGACGACCTGCACTGTGAAAAAA 1020
QY 1021 GACAGCCTTAAACGAAAACCATTAATTCGGAACCTGACGAAAGACTTAATTTGCTTGG 1080
DB 1021 GACAGCCTTAAACGAAAACCATTAATTCGGAACCTGACGAAAGACTTAATTTGCTTGG 1080

```

```

QY 1081 AAAGTAGAAGACGGAAGAACTGACGTTTGATATGATCATAGTACAAACTAAATCTAAA 1140
DB 1081 AAAGTAGAAGACGGAAGAACTGACGTTTGATATGATCATAGTACAAACTAAATCTAAA 1140

RESULT 3
ADV65734
ID ADV65734 standard; DNA; 1140 BP.
XX
XX
AC ADV65734;
XX
XX
DT 10-FEB-2005 (first entry)
XX
XX
DE B. subtilis poly-gamma-glutamic acid synthase A gene SEQ ID NO:3.
XX
XX
KW ds; poly-gamma-glutamic acid synthase A; pgsA; gene expression;
XX
XX
KW antibiotic; antimicrobial; fungicide; cyclostatic.
OS Bacillus subtilis.
XX
XX
PN KR2004034780-A.
XX
XX
PD 29-APR-2004.
XX
XX
PF 17-OCT-2002; 2002KR-00063379.
XX
XX
PR 17-OCT-2002; 2002KR-00063379.
XX
XX
PA (BIOL-) BIOLEADERS CORP.
XX
XX
PA (UYCH-) UNIV CHUNGSUN CO LTD.
XX
XX
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
XX
PI Boo HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
XX
XX
PI Park YG, Sung MH;
XX
XX
DR WPI; 2004-577380/56.
XX
XX
PT Method for surface expression of peptides p5 and anal3 using pgs bca
PT gene.
XX
XX
PS Disclosure; SEQ ID NO 3; 25pp; Korean.
XX
XX
XX
CC The invention relates to a novel method for surface expression of
CC peptides p5 and anal3 using a poly-gamma-glutamic acid synthase (pgs) Bca
CC gene, thereby removing a purification process of peptides p5 and anal3,
CC and using lactic acid bacteria for the surface expression, so that
CC peptide antibiotics can be cheaply and stably mass-produced. An
CC expression vector pHCpLb:pgsA-p5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide p5 encoded by the nucleotide sequence set forth
CC in ADV65735, or to the peptide anal3 encoded by the nucleotide sequence
CC set forth in ADV65737. The present sequence represents the B. subtilis
CC poly-gamma-glutamic acid synthase A gene used in the invention.
XX
XX
SQ Sequence 1140 BP; 381 A; 235 C; 265 G; 259 T; 0 U; 0 Other;

Query Match 100.0%; Score 1140; DB 13; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAAAGAACTGAGCTTTGATGAAAAGCTCTTAAAGCTGACAAAACAGAAAAAAG 60
DB 1 ATGAAAAAAGAACTGAGCTTTGATGAAAAGCTCTTAAAGCTGACAAAACAGAAAAAAG 60
QY 61 AAAACCAATPAGCAGTATTTATGTCATTCGATCGTGTGCTTATGTTGCTTTC 120
DB 61 AAAACCAATPAGCAGTATTTATGTCATTCGATTCGATCGTGTGCTTATGTTGCTTTC 120
QY 121 ATGTGGCGGGGAAAAGCGGAAAACGCCGAAGTCAAAAGCTATTTGACGACGACTCTCA 180
DB 121 ATGTGGCGGGGAAAAGCGGAAAACGCCGAAGTCAAAAGCTATTTGACGACGACTCTCA 180

```

```
QY 181 GCGTCATTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 240
DB 181 GCGTCATTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 240
QY 241 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATGACA 300
DB 241 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATGACA 300
QY 301 GGAACCTTTGAAAAACCCGGTAACTTATCAAAAAGATTATAACAAGCAGATTAAGAGATT 360
DB 301 GGAACCTTTGAAAAACCCGGTAACTTATCAAAAAGATTATAACAAGCAGATTAAGAGATT 360
QY 361 CATCTGCGAGAGATTAAGGAATCAGTAAAGTCTTGAAGAGATATGAAATTTCAAGGTTCTC 420
DB 361 CATCTGCGAGAGATTAAGGAATCAGTAAAGTCTTGAAGAGATATGAAATTTCAAGGTTCTC 420
QY 421 AACAGCGCAACCAACGCAATGATTAACGGCGTTACAGGGCATGAAAGATACGCTTGA 480
DB 421 AACAGCGCAACCAACGCAATGATTAACGGCGTTACAGGGCATGAAAGATACGCTTGA 480
QY 481 GAATTGGCAAGCAAAACCTTGATATCGTTGAGCGGGATACAGTTAAAGTATCGCAAA 540
DB 481 GAATTGGCAAGCAAAACCTTGATATCGTTGAGCGGGATACAGTTAAAGTATCGCAAA 540
QY 541 AAGAAATTTTGTACAGAAAGTCAACGGGGTAAAGTATGCAAGCTTGGCTTTACCGAT 600
DB 541 AAGAAATTTTGTACAGAAAGTCAACGGGGTAAAGTATGCAAGCTTGGCTTTACCGAT 600
QY 601 GTGTCCGGGAAAGTCTTTCGCGCTAAAAAGATATACCGCGGGCTGCTCGCGCAGATCT 660
DB 601 GTGTCCGGGAAAGTCTTTCGCGCTAAAAAGATATACCGCGGGCTGCTCGCGCAGATCT 660
QY 661 GAATCTTCATCCTTATGATTTCAAGAGGAAAAACATGCTGACATTTGTTGTGAG 720
DB 661 GAATCTTCATCCTTATGATTTCAAGAGGAAAAACATGCTGACATTTGTTGTGAG 720
QY 721 TCACAGCTGGGGCAGAGATGATGACAAATATCCAAAGCAGCGGCGAGCTTGACAGA 780
DB 721 TCACAGCTGGGGCAGAGATGATGACAAATATCCAAAGCAGCGGCGAGCTTGACAGA 780
QY 781 GGCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCCGACAGTCTTTAGAACG 840
DB 781 GGCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCCGACAGTCTTTAGAACG 840
QY 841 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCCTTCGCAACTTTGTTGACCAA 900
DB 841 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCCTTCGCAACTTTGTTGACCAA 900
QY 901 GCGTGGAGAGAAACAAGACAGTGCATGCTTCAAGTATCACTGTAAGAAAAATGGAACA 960
DB 901 GCGTGGAGAGAAACAAGACAGTGCATGCTTCAAGTATCACTGTAAGAAAAATGGAACA 960
QY 961 GGGCGCTTTGAAGTACACCGATCATATCCATGGAAGCAGACCTGCACTGTGAAAAA 1020
DB 961 GGGCGCTTTGAAGTACACCGATCATATCCATGGAAGCAGACCTGCACTGTGAAAAA 1020
QY 1021 GACAGCCTTTAAACAGAAAAACATTAATGCGCAACTGACGAAAGACTTAATTTCCCTGG 1080
DB 1021 GACAGCCTTTAAACAGAAAAACATTAATGCGCAACTGACGAAAGACTTAATTTCCCTGG 1080
QY 1081 AAAGTGAAGAGGAAAGCTGACGTTGATATTTGATGATGATGATGATGATGATGATGAT 1140
DB 1081 AAAGTGAAGAGGAAAGCTGACGTTGATATTTGATGATGATGATGATGATGATGATGAT 1140
```

RESULT 4
ADA44796 standard; DNA; 6536 BP.
AC ADA44796;
XX
XX
DT 20-NOV-2003 (first entry)

```
XX DE Microbial cell surface expression vector pGNBCA, SEQ ID NO:6.  
XX KW Cell surface; expression vector; microbial;  
KW poly-gamma-glutamate synthetase; Bacillus subtilis pgsBCA complex;  
KW surface expression; Gram-positive bacterium; Gram-negative bacterium;  
KW enzyme; antigen; antibody; attachment protein; adsorption protein;  
KW vaccine; pGNBCA; cyclic; circular; ds.  
XX  
OS Synthetic.  
OS Bacillus subtilis; variety chungkookjang.  
PN MO2003014360-A1.  
XX  
XX PD 20-FEB-2003.  
XX  
XX PF 09-AUG-2002; 2002WO-KR001522.  
XX  
XX PF 10-AUG-2001; 2001KR-00048373.  
XX  
XX PA (BIOL-) BIOTRADERS CORP.  
XX (MDM-) MD LAB CO LTD.  
XX  
XX PI Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashiuchi M;  
XX WPI; 2003-256589/25.  
XX  
XX PS New expression vector containing gene(s) that encode a poly-gamma-  
PT glutamate synthetase complex, useful for producing proteins (e.g.  
PT vaccines or enzymes) on the microbial surface of Gram-positive and/or  
PT Gram-negative bacteria.  
XX  
XX Example 1; Page 113-120; 122pp; English.  
XX  
XX The invention relates to a vector for expression of a target protein on a  
CC microbial cell surface. The vector of the invention comprises either one  
CC or more than two genes of the Bacillus subtilis poly-gamma-glutamate  
CC synthetase complex (pgsBCA) to facilitate microbial surface expression of  
CC the target protein. The pgsBCA gene complex comprises the pgsB, pgsB and  
CC pgsA genes and is normally expressed in the outer membrane of Bacillus  
CC subtilis. The vector can be transformed into either Gram-positive or Gram  
CC -negative bacteria (e.g., Escherichia coli), and can be used for the  
CC surface expression of various proteins of interest such as enzymes,  
CC antigens, antibodies, attachment proteins or adsorption proteins.  
CC Proteins recombinantly produced using the vector of the invention can be  
CC used as, for example, vaccines or enzymes. The present sequence  
CC represents a microbial cell surface expression vector of the invention,  
XX pGNBCA.  
XX  
SQ Sequence 6536 BP; 1855 A; 1464 C; 1567 G; 1649 T; 0 U; 1 Other;  
Query Match 100.0%; Score 1140; DB 10; Length 6536;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTGTAAGCTGACAAAAAGCAAAAAAAG 60  
DB 1891 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTGTAAGCTGACAAAAAGCAAAAAAAG 1950  
QY 61 AAAACCAATGAAGCAGTATTTATGCAATTCGATCGATGCTTTTGTCTTATGTTGCTTTC 120  
DB 1951 AAAACCAATGAAGCAGTATTTATGCAATTCGATCGATGCTTTTGTCTTATGTTGCTTTC 2010  
QY 121 ATGTGGCGGGGAAAAAGCGGAAACGCCCAAGTCAAAACGTTTCTGACGAGCTCTCA 180  
DB 2011 ATGTGGCGGGGAAAAAGCGGAAACGCCCAAGTCAAAACGTTTCTGACGAGCTCTCA 2070  
QY 181 GCGTCATTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 240  
DB 2071 GCGTCATTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 2150  
QY 241 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATGACA 300
```

Db 2131 GGGGAGACATATTTTCAATATGTTGACGATCTTTAGAGCTGGATATATGACA 2130
QY 301 GGAACCTTTGAAACCCGGTAACCTATCAAAAGATTATTAACAAGCATTAAGAGATT 360
Db 2191 GGAACCTTTGAAACCCGGTAACCTATCAAAAGATTATTAACAAGCATTAAGAGATT 2250
QY 361 CATCTGCAGAGATTAAGGAATCACTGAAAGCTTGAAGGAATTAAGGAATTAAGGAAT 420
Db 2251 CATCTGCAGAGATTAAGGAATCACTGAAAGCTTGAAGGAATTAAGGAATTAAGGAAT 2310
QY 421 AACGCGCCCAACACCAAGCAATGATTAAGGCGCTTCAAGGCGATGAAGATACGCTTGA 480
Db 2311 AACGCGCCCAACACCAAGCAATGATTAAGGCGCTTCAAGGCGATGAAGATACGCTTGA 2370
QY 481 GAATTTGGAGAACCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAGTATCGAAA 540
Db 2371 GAATTTGGAGAACCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAGTATCGAAA 2430
QY 541 AAGAAAATTTGATACAGAAAGTCAACGGGGTAAAGATTTGCAAGCTTGGCTTACCGAT 600
Db 2431 AAGAAAATTTGATACAGAAAGTCAACGGGGTAAAGATTTGCAAGCTTGGCTTACCGAT 2490
QY 601 GTGTCGGGGAAGGTTTGGCGGCTAAAAAGAAATACGCGGCGCTGCTGCGCGAGATCT 660
Db 2491 GTGTCGGGGAAGGTTTGGCGGCTAAAAAGAAATACGCGGCGCTGCTGCGCGAGATCT 2550
QY 661 GAAATCTTCACTCCCTATGATTTCAAGAGGAAAAAATCATGCTAGACATTTGTTGCGAG 720
Db 2551 GAAATCTTCACTCCCTATGATTTCAAGAGGAAAAAATCATGCTAGACATTTGTTGCGAG 2610
QY 721 TCACACTGGGCGCAAGATGATGACATGATCCAAACGACCGGCGACCGGCGACGCTTGAAGA 780
Db 2611 TCACACTGGGCGCAAGATGATGACATGATCCAAACGACCGGCGACCGGCGACGCTTGAAGA 2670
QY 781 GGCATGTCGATGCGGAGGCTGACATCATGCTGCGGCGATCATCCGCACTTTAGAACCG 840
Db 2671 GGCATGTCGATGCGGAGGCTGACATCATGCTGCGGCGATCATCCGCACTTTAGAACCG 2730
QY 841 ATTGAAGTATATACGGAACCGTCAATTTCTACAGCTTGGCACTTTGCTTTGACCA 900
Db 2731 ATTGAAGTATATACGGAACCGTCAATTTCTACAGCTTGGCACTTTGCTTTGACCA 2790
QY 901 GGCTGACGAGAACAGAGACAGTGCATGCTGCTTCAATATCACTGAAAGAAATGAGACA 960
Db 2791 GGCTGACGAGAACAGAGACAGTGCATGCTGCTTCAATATCACTGAAAGAAATGAGACA 2850
QY 961 GGGCGCTTTGAAGTACACCGATCCATATCCATGAAAGGACACCTGCACTTGAAGAAA 1020
Db 2851 GGGCGCTTTGAAGTACACCGATCCATATCCATGAAAGGACACCTGCACTTGAAGAAA 2910
QY 1021 GAGAGCTTAAACAGAAACCACTTATTCGGAAGTGAAGAAAGACTTATTTGCTTGG 1080
Db 2911 GAGAGCTTAAACAGAAACCACTTATTCGGAAGTGAAGAAAGACTTATTTGCTTGG 2970
QY 1081 AAGTAGAAGACGAAAGTGAAGTGAATTTGATGATGATGATGATGATGATGATGATGAT 1140
Db 2971 AAGTAGAAGACGAAAGTGAAGTGAATTTGATGATGATGATGATGATGATGATGATGAT 3030

RESULT 5
AAF82257
ID AAF82257 standard; DNA; 1143 BP.

AAF82257;

21-JUN-2001 (first entry)

Bacillus subtilis IFO 3336 DNA encoding a PGA synthesizing enzyme.

Bacterial; poly-gamma-L-glutamic acid; L-PGA;

KM poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; ds.

Bacillus subtilis.

XX Key Location/Qualifiers
FH CDS 1..1143
FT /*tag= a
FT CDS
PN JP2001017182-A.
XX 23-JAN-2001.
XX 09-JUL-1999; 99JP-00196335.
XX 09-JUL-1999; 99JP-00196335.
XX 09-JUL-1999; 99JP-00196335.
XX (NAGS) NAGASE SANGYO KK.
XX WPI; 2001-285408/30.
XX DR P-PSDB; MAB74026.
XX DR P-PSDB; MAB74026.
XX PT New nucleic acid encoding a glutamate racemase enzyme useful for the
XX preparation of poly-gamma-glutamic acid.
XX PS Claim 4; Page 13-15; 17pp; Japanese.
XX CC The present sequence encodes an enzyme which is useful in the production
XX of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L-glutamic acid
XX (PGA). A plasmid comprising the present sequence may be used to transform
XX CC Escherichia coli. The transformants express the enzyme and PGA is
XX CC produced in the culture
XX SQ Sequence 1143 BP; 385 A; 233 C; 264 G; 261 T; 0 U; 0 Other;
Query Match 99.3%; Score 1132; DB 4; Length 1143;
Best Local Similarity 99.6%; Pred. No. 0; Mismatches 5; Indels 0; Gaps 0;
Matches 1135; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGAAAAAAGAACTGAGCTTTCAATGAAAGCTGTAAGCTGACAAAACAGCAAAAAAG 60
Db 1 ATGAAAAAAGAACTGAGCTTTCAATGAAAGCTGTAAGCTGACAAAACAGCAAAAAAG 60
QY 61 AAAACCAATAGACAGATTAATTTATGCGATTCGATCGTTTGTCTTATGTCCTTTC 120
Db 61 AAAACCAATAGACAGATTAATTTATGCGATTCGATTCGATCGTTTGTCTTATGTCCTTTC 120
QY 121 ATGTGGGCGGGAAGGCGGAAACGCGAAGGTCAAAAGTATTCGACGACGCTACTTCA 180
Db 121 ATGTGGGCGGGAAGGCGGAAACGCGAAGGTCAAAAGTATTCGACGACGCTACTTCA 180
QY 181 GCTCATTTGTAAGGAGATTAATGATGGAAGCTATGTTGAAAAAGTAACGAGCAAAA 240
Db 181 GCTCATTTGTAAGGAGATTAATGATGGAAGCTATGTTGAAAAAGTAACGAGCAAAA 240
QY 241 GGGGAGACAGATTTTCAATATGTTGAACGATCTTTAGAGCTCGGATTAATGAGA 300
Db 241 GGGGAGACAGATTTTCAATATGTTGAACGATCTTTAGAGCTCGGATTAATGAGA 300
QY 301 GGAACCTTTGAAACCCGGTAACCTATCAAAAGATTATTAACAAGCATTAAGAGATT 360
Db 301 GGAACCTTTGAAACCCGGTAACCTATCAAAAGATTATTAACAAGCATTAAGAGATT 360
QY 361 CATCTGCAGAGATTAAGGAATCACTGAAAGCTTGAAGGAATTAAGGAATTAAGGAAT 420
Db 361 CATCTGCAGAGATTAAGGAATCACTGAAAGCTTGAAGGAATTAAGGAATTAAGGAAT 420
QY 421 AACGCGCCCAACACCAAGCAATGATTAAGGCGCTTCAAGGCGATGAAGATACGCTTGA 480
Db 421 AACGCGCCCAACACCAAGCAATGATTAAGGCGCTTCAAGGCGATGAAGATACGCTTGA 480
QY 481 GAATTTGGAGAACCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAGTATCGAAA 540
Db 481 GAATTTGGAGAACCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAGTATCGAAA 540
QY 541 AAGAAAATTTGATACAGAAAGTCAACGGGGTAAAGATTTGCAAGCTTGGCTTACCGAT 600
Db 541 AAGAAAATTTGATACAGAAAGTCAACGGGGTAAAGATTTGCAAGCTTGGCTTACCGAT 600

```
Db 541 AAGAAATTTTGTGATCAAAAAGTCAACGGGGTAACGATTTGGACGCTTGCTTACCGAT 600
Qy GTGTCGGGAAAGGTTTCGGGCTTAAAGAAATCGCCGGGCGTCTGCGGAGATCT 660
Db 601 GTGTCGGGAAAGGTTTCGGGCTTAAAGAAATCGCCGGGCGTCTGCGGAGATCT 660
Qy 661 GAAATCTTCATCCCTATGATTTTCAGAAAGCGAAAAACATGCTGACATTTGTTGTGCAG 720
Db 661 GAAATTTTCATCCCTATGATTTTCAGAAAGCGAAAAACATGCTGACATTTGTTGTGCAG 720
Qy 721 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCGACGCTTGCAAGA 780
Db 721 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCGACGCTTGCAAGA 780
Qy 781 GCCATGCTGATGGGGAGCTGACATCATCGTCGGGCGATCATCCGCACTGTTGAACCG 840
Db 781 GCCATGCTGATGGGGAGCTGACATCATCGTCGGGCGATCATCCGCACTGTTGAACCG 840
Qy 841 ATTGAAGTATATTAACGAAACCGTCATTTTCTACAGCCCTCGGCACTTTGCTTGAACCA 900
Db 841 ATTGAAGTATATTAACGAAACCGTCATTTTCTACAGCCCTCGGCACTTTGCTTGAACCA 900
Qy 901 GCGTGGACGAGAACAGAGACAGTGCACCTGCTTCAGTATCACTGAAAGAAATGGAACA 960
Db 901 GCGTGGACGAGAACAGAGACAGTGCACCTGCTTCAGTATCACTGAAAGAAATGGAACA 960
Qy 961 GCGGCGCTTTGAAGTACACCGATCGATTCATGTAAGGAGACCTTGCACTTGTAAGAAA 1020
Db 961 GCGGCGCTTTGAAGTACACCGATCGATTCATGTAAGGAGACCTTGCACTTGTAAGAAA 1020
Qy 1021 GACAGCCTTTAAACAGAAACCATTAATTCGCGACATGACGAAAGACTTAATTCGCTTGG 1080
Db 1021 GACAGCCTTTAAACAGAAACCATTAATTCGCGACATGACGAAAGACTTAATTCGCTTGG 1080
Qy 1081 AAAGTAGAAGACGGAACCTGACGTTTGATGATGATGATGACAAACTAAATCTAAA 1140
Db 1081 AAAGTAGAAGACGGAACCTGACGTTTGATGATGATGATGACAAACTAAATCTAAA 1140

RESULT 6
AAF82254
ID AAF82254 standard; DNA; 3045 BP.
XX AC AAF82254;
XX DT 21-JUN-2001 (first entry)
XX DE Bacillus subtilis IFO 3336 DNA encoding glutamate racemase enzyme.
XX KW Bacterial; poly-gamma-L-glutamic acid; L-PGA;
XX KW poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; ds.
XX OS Bacillus subtilis.
XX Key Location/Qualifiers
FH CDS 19..1200
FT /*tag= a
FT CDS 1215..1264
FT /*tag= b
FT CDS 1683..2825
FT /*tag= c
XX JP2001017182-A.
XX 23-JAN-2001.
XX PD 09-JUL-1999; 99JP-00196335.
XX PF 09-JUL-1999; 99JP-00196335.
XX PR 09-JUL-1999; 99JP-00196335.
XX PA (NAGS ) NAGASE SANGYO KK.
XX DR WPI; 2001-285408/30.
```

```
DR P-PSDB; AAB74024, AAB74025, AAB74026.
XX New nucleic acid encoding a glutamate racemase enzyme useful for the
PT preparation of poly-gamma-glutamic acid.
XX Claim 1; Page 7-11; 17pp; Japanese.
XX The present sequence encodes a glutamate racemase enzyme which is useful
CC in the production of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L
CC -glutamic acid (PGA). A plasmid comprising the present sequence may be
CC used to transform Escherichia coli. The transformants express the enzyme
CC and PGA is produced in the culture
XX
SQ Sequence 3045 BP; 1000 A; 593 C; 696 G; 756 T; 0 U; 0 Other;
Query Match 99.3%; Score 1132; DB 4; Length 3045;
Beet Local Similarity 99.6%; Pred. No. 0;
Matches 1135; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ATGAAAAAAGACTGACCTTTCATGAAAAGCTGTAAGCTGACAAAAACAGCAAAAAAG 60
Db 1683 ATGAAAAAAGACTGACCTTTCATGAAAAGCTGTAAGCTGACAAAAACAGCAAAAAAG 1742
Qy 61 AAAACCAATTAAGCAGTATTTTATTCGCTTCGATCGATGCTTTTGTCTTATGTTGCTTTC 120
Db 1743 AAAACCAATTAAGCAGTATTTTATTCGCTTCGATCGATGCTTTTGTCTTATGTTGCTTTC 1802
Qy 121 ATGTGGCGGGGAAAGGCGGAAACGCGAAGTCAAAACGATTCGTGACGATCTCTCA 180
Db 1803 ATGTGGCGGGGAAAGGCGGAAACGCGAAGTCAAAACGATTCGTGACGATCTCTCA 1862
Qy 181 GCCTCATTTGAGCGATATTAATGATGAGGACGCTATGTTGAAAAAGTAAAGCAAAAA 240
Db 1863 GCCTCATTTGAGCGATATTAATGATGAGGACGCTATGTTGAAAAAGTAAAGCAAAAA 1922
Qy 241 GGGGCAACAGTATTTTCAATATGTTGAAACCGATCTTTAGAGCTCGGATATGTAACA 300
Db 1923 GGGGCAACAGTATTTTCAATATGTTGAAACCGATCTTTAGAGCTCGGATATGTAACA 1982
Qy 301 GGAACCTTTGAAAAACCGGTTAACCTATCAAAAGAAATTAATAACAGAGATTAAGAAATT 360
Db 1983 GGAACCTTTGAAAAACCGGTTAACCTATCAAAAGAAATTAATAACAGAGATTAAGAAATT 2042
Qy 361 CATCTGACAGAGATTAAGAAATCAGTAAAGTCTTGAAGATATGAAATTTACCGTTCTC 420
Db 2043 CATCTGACAGAGATTAAGAAATCAGTAAAGTCTTGAAGATATGAAATTTACCGTTCTC 2102
Qy 421 AACAGCGCAACCAACAGCGAATGATTAACGCGTTCAAGGCAATGAAGATACGCTTGA 480
Db 2103 AACAGCGCAACCAACAGCGAATGATTAACGCGTTCAAGGCAATGAAGATACGCTTGA 2162
Qy 481 GAAATTTGCAAGCAAAACCTTGATATGCTTGAGACGGGATPACGCTTAAGATGAGCGAAA 540
Db 2163 GAAATTTGCAAGCAAAACCTTGATATGCTTGAGACGGGATPACGCTTAAGATGAGCGAAA 2222
Qy 541 AAGAAAAATTTGCTACAGAAAAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 600
Db 2223 AAGAAAAATTTGCTACAGAAAAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 2282
Qy 601 GTGTCCGGGAAAGGTTTCGCGGCTTAAAGAAATACGCGGCGTGCTGCCGAGATCTT 660
Db 2283 GTGTCCGGGAAAGGTTTCGCGGCTTAAAGAAATACGCGGCGTGCTGCCGAGATCTT 2342
Qy 661 GAAATCTTCATCCCTATGATTTTCAGAAAGCGAAAAACATGCTGACATTTGTTGTGCAG 720
Db 2343 GAAATCTTCATCCCTATGATTTTCAGAAAGCGAAAAACATGCTGACATTTGTTGTGCAG 2402
Qy 721 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCGACGCTTGCAAGA 780
Db 2403 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCGACGCTTGCAAGA 2462
Qy 781 GCCATGCTGATGGGGAGCTGACATCATCGTCGGGCGATCATCCGCACTGTTGAACCG 840
```



```
D 2463 GCCATGTCGATGCGGAGCTGATCATCGTCGCCATCATCCGACGCTTGAACCG 2522
Q 841 ATTGAAGTATTAACGGAACCGTCATTTTCTACAGCCTCGCACTTGTCTTGACCA 900
D 2523 ATTGAAGTATTAACGGAACCGTCATTTTCTACAGCCTCGCACTTGTCTTGACCA 2582
Q 901 GCGTGAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 960
D 2583 GCGTGAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2642
Q 961 GCGGCTTGAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1020
D 2643 GCGGCTTGAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2702
Q 1021 GACAGCCTTAAACAGAAAACCTTATTCGCGAAGTACGAAAAGCTTAAATTCCTGG 1080
D 2703 GACAGCCTTAAACAGAAAACCTTATTCGCGAAGTACGAAAAGCTTAAATTCCTGG 2762
Q 1081 AAGTGAAGACGAAAACCTGATTCGTTGATTCATTCATTCATTCATTCATTCATTC 1140
D 2763 AAGTGAAGACGAAAACCTGATTCGTTGATTCATTCATTCATTCATTCATTCATTC 2822
```

RESULT 7

ABK74454
ID ABK74454 standard; DNA; 696 BP.

AC ABK74454;

XX 13-AUG-2002 (first entry)

DE Bacillus licheniformis genomic sequence tag (GST) #1745.

KM Differential gene expression; genomic sequenced tag: GST;
altered culture condition; environmental stress;

XX physiological provocation; ds.

OS Bacillus licheniformis.

PN WO200229113-A2.

XX 11-APR-2002.

XX 05-OCT-2001; 2001WO-US031437.

XX 06-OCT-2000; 2000US-00680598.

XX 27-MAR-2001; 2001US-0279526P.

XX (NOVO) NOVOZYMES BIOTECH INC.

XX (NOVO) NOVOZYMES AS.

XX Berka R, Clausen IG;

XX WPI; 2002-41684/44.

XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
array.

XX Claim 4; SEQ ID NO 1745; 200pp; English.

CC The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labeled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and

CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at http://wipo.int/pub/published_pct_sequences

CC Sequence 696 BP; 226 A; 151 C; 173 G; 146 T; 0 U; 0 Other;

SQ Query Match 29.2%; Score 332.4; DB 6; Length 696;

Best Local Similarity 68.4%; Pred. No. 7.1e-85;

Matches 475; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

```
Q 444 GGATTACGCGCTTCAGGCGATGAAAGATACGCTTGAGAAATTTGGAGCAAAACCTTGA 503
D 1 GGACTACGCGGGAAGACGCTTGAAGATACGCTCAATTAATTTCAATGAGAAATCTGA 60
Q 504 TATCGTGAAGCGGATACAGCTTAAGTATGCGCAAAAGAAATTTGCTACCGAAGT 563
D 61 GCTGTGCGAGCAGGAAATTAATCTTGAAAGCGCAAGACGATCTATCAGAAATGT 120
Q 564 CAACGGGTTACGATTCGAAAGCTTGCTTACGATGTCGCGGAAAGTTTGCAGC 623
D 121 GAACGGGTTAAATTTGACGCTGCTGTTTACAGACGCTTACAAAGAACTTACAGC 180
Q 624 TAAAGAATACGCGCGGCTGCTGCGCGAGATCTGAAATCTTCATCCCTATGATTC 683
D 181 CAAAAGAAACAGAGCGGAGAGTGTCCGCTCAGTCCGAAATCTTATTCATGATTC 240
Q 684 AGAACGAAAAAATGCTGATTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 743
D 241 GGAACATGCAAAAAAAGCGGATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Q 744 CATGATCCAAACGACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 803
D 301 CATGATCCAAACGACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Q 804 CATGATCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 863
D 361 TGTGATCATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Q 864 CATTTTCTACAGCCTCGGCACTTTGTCTTGACCAAGGCTGACGAGCAAGACAG 923
D 421 GATTTTCTACAGCCTCGGCACTTTGTATTTGATGACGAGGCTGATCAAGAACAGG 480
Q 924 TGCATGCTTCAATGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 983
D 481 CGCGCTTGTACAAATCAATTAATGATGACGCAAGGCGCTTGTGAGTACGCTCTT 540
Q 984 CGATATCCATGAAGACGACCTGCACT---GTGAAAAAAGACAGCTTAAACAGAAAC 1040
D 541 CAACATTCGCAAGACGACCGCAAGCTTTAGCAAGACGACTTTTAAAGAAAGAC 600
Q 1041 CATTTATTCGCAAGACGCAAGAACTTAAATTTGCTTGAAGTGAAGAGGAAACT 1100
D 601 GATCTTCGCTCAATGACAAAGAAACAAACCTGCACTGGAAGAAAGAAAGGAAATTT 660
Q 1101 GACGTTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1134
D 661 AACGTTGAAGTCATGCGGACAAAGCTGAA 694
```

RESULT 8

ABK74486
ID ABK74486 standard; DNA; 488 BP.

XX ABK74486;

XX 13-AUG-2002 (first entry)


```

XX DE Bacillus licheniformis genomic sequence tag (GST) #1777.
XX XX
XX XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX XX
XX OS Bacillus licheniformis.
XX WO200229113-A2.
XX XX
XX PD 11-APR-2002.
XX PF 05-OCT-2001; 2001WO-US031437.
XX PR 06-OCT-2000; 2000US-00680598.
XX PR 27-MAR-2001; 2001US-0279526P.
XX PA (NOVO ) NOVOZYMES BIOTECH INC.
XX PA (NOVO ) NOVOZYMES AS.
XX PI Berka R, Clausen IG;
XX DR WPI; 2002-416684/44.
XX PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX CS Claim 4; SEQ ID NO 1777; 200pp; English.

```

CC The invention describes a method of monitoring differential expression of
CC genes in a first *Bacillus* cell relative to expression of the genes in
CC other *Bacillus* cells, comprising hybridising labelled nucleic acid probes
CC isolated from *Bacillus* cells to a substrate containing array of *Bacillus*
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first *Bacillus* cell relative to expression of the same genes
CC in one or more second *Bacillus* cells. The method is useful for monitoring
CC global expression of several genes from a *Bacillus* cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which *Bacillus* cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at http://www.wipo.int/pub/published_pct_sequences
XX

Sequence 468 BP; 178 A; 89 C; 105 G; 116 T; 0 U; 0 Other;

Query Match	19.8%	Score 225.6;	DB 6;	Length 488;
Best Local Similarity	67.8%	Pred. No. 3.3e-54;		
Matches 331; Conservative	0;	Mismatches 154;	Indels 3;	Gaps 1;

QY	1	ATGAAAAAAAAAGATGAGCTTTTCATGAAAAGCTGCTAAAGTGACMAAACGCAAAAAAAG	60
Db	1	ATGAAAAAAAACAACTGAACTTTCAGGAAAACTGCTAAAGTTGACGAAGCGAGAAAAAG	60
QY	61	AAACCAATPAACGATTTTATTTGACATTCGATGGTTTTTTCCTATAGTTCGCTTC	120
Db	61	AAAAACAACAGACGCTTTATCGATATGCCGTAATTTTCGTTAAAGTTGCTTT	120
QY	121	ATGTGGCGCGGAAAAAGCGAAAAAGC---CGAAGGTCAAAAAGCTATTTCGACGAGTACTC	177
Db	121	ACTTGGGTGGGAAGCGCCAAAAGCTCTTCGCAATGACCAAAAAAGAAAGATGCCAAGTT	180
QY	178	TCAGGCTCATTTGTAGCGCATATTATGATGGACGCTATGTTGAAAAAGTAACGAGCA	237

Db	181	ACAGCTACTTTTGTGTGGGATATCATGATGGGAAAGAAACGTAGAAAAAGTGACAAACTTG	240
Qy	238	AAAGGGGCAGACGATATTTTTCATATGTTGAACCGATCTTTAGAGCTCGGATATATGA	297
Db	241	CACGGTTTCGGAAAGTCTCTTCAAAAATGTGAAGCCGTACTTTAATGTGTCAGATTTATC	300
Qy	298	GCAGGAACTTTGAAAACCCGGTAACTCAATCAAAAAGATTTAAACAAGCAGATTAAGAG	357
Db	301	ACGGAACCTTTGAAAAACCTGTAACTCAATGCAAGACGATCAAGAGCCGAAAAGAAC	360
Qy	358	ATTCACTCTCAGACGATATAGGAATCAGTGAAGTCTTGAAGATATGAATTTACCGTT	417
Db	361	ATCATCTTCGAAACGATCATAGAAATCAATCGAAACATTTGAAAAAGCTGAACCTTACGGTA	420
Qy	418	CTCAACAGCGCCAAACAACCGCAATGGAATTAACGCGTTCAAGGCATGAAAGATACGTT	477
Db	421	CTGAATTTTTCGCAACACCATCGATGACTACGGGGAGACGCGTTTGAAGGATACGCTC	480
Qy	478	GGAGAAATT	485
Db	481	AATGAGTT	488

RESULT 9
ADB06115
ID ADB06115 standard; DNA; 1212 BP.

AC ADB06115;

DT 20-NOV-2003 (first entry)

DE A10iococcus otitis antigenic protein encoding DNA SEQ ID NO:55.

Alloccoccus otitidis; antigenic protein; immunogenic; immunisation, KW
KW
gene therapy; Gram-positive bacterium; infection; gene; ds.

Aloiococcus otitis.

PN WO2003048304-A2

PD 12-JUN-2003

PF 25-NOV-2002; 2002WO-US036123.

PR 29-NOV-2001; 2001US-0333777P.

XX XX

XXXXXX

[illegible]

DR P-PSDB; ADB06116.

PT New Alliococcus otr

PT effects during drug use

PS Claim 7; SEQ ID NO 5

CC The present invention

CC *Alloicoccus otitidis*

CC expression vector co

CC host cell, transfect

composition comprises

CC expression vector;
CC polypeptide of (1)

CC The present invention describes an isolated polynucleotide (1) of
CC *Alloicoccus oltidis* genomic DNA, which encodes an antigenic protein,
CC *Alloicoccus oltidis* is a Gram-positive bacterium. Also described: (1)
CC an isolated polynucleotide that is encoded by the polynucleotide (1); (2)
CC an expression vector comprising the novel isolated polynucleotide (1), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array

of the polypeptides of (1), their biological equivalent or fragment; (8) immunising against *Allotiococcus ostitidis* by administering to a host the immunogenic composition; (9) detecting and/or identifying *Allotiococcus ostitidis* in the biological sample; (10) a kit comprising a container containing the novel polynucleotide, its degenerate variant or fragment, or the antibody of (4); and (11) producing a polypeptide by culturing the genetically engineered host cell under conditions suitable to produce the polypeptide from the culture. (1) can be used in gene therapy. The polynucleotides, polypeptides, antibodies and compositions of the present invention can be used for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials. The polynucleotides are useful for expressing and detecting *Allotiococcus ostitidis*. The present sequence encodes an *Allotiococcus ostitidis* antigen protein from the present invention.

Sequence 1212 BP; 353 A; 248 C; 259 G; 352 T; 0 U; 0 Other;

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
Best Local Similarity 46.7%; Pred. No. 7.8e-17;
Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

```

184 TCATTGTGAGGAGATTTATGATGAGGACCGCTATGTTGAAAAAGTAAACGAGCAAAAAGGG 243
238 TCATTGTGAGGAGATTTATGATGAGGACCGCTATGTTGAAAAAGTAAACGAGCAAAAAGGG 297
244 GCAGACAGATTTTTCATATGTTGAACCGATCTTTAGAGCTCGGATTTATGACAGCA 303
298 TAGAGAGATTTTTCATATGTTGAACCGATCTTTAGAGCTCGGATTTATGACAGCA 357
304 AACCTTGAACACCGGTTAATCTTCAAAAAGATTA-----ACAGCA 348
358 AATCTTGAATCTGCGCTACATGACATCGAATCTTACGTTAAGCCGCTTCAAGTTTA 417
349 GATTAAGAGATTCATCTGACAGCAATTAAGAAATCAGTAAAGTCTTGAAGATTAAT 408
418 AGTGAGAGATTTTATTTGATTTCTTCCATTTGACTGACATCAAGCAAGAGTGGC 477
409 TTCACGGTTCTCAACAGCGCAACCAACGCAATGATTAACGGCTTCAAGGCAATGAAA 468
478 ATTAGACCTGATTTCCATGCGCAACCAATACCGGCAACATGCGGATGATGTT 537
469 GATACCTTGAAGATTTTGAAGCAAAACCTGATATGTTGACCGGATACAGCTTA 528
538 GACGAGATGAAATTTTAAGAGATGAAATGATTAATTTGAAATGCGCATGATCGA 597
529 AGTATGAGCAAAAAGATTTGTTACAGAAAGTCAACGGGTTAAGATTCAGCTT 588
598 GCAGAAAGC--TGCCAGCTTATCAATTTGTCGATATATTTGACTACCTCATTTT 654
589 GACCTTACCGATGTGTCGCGGAAAGTTTGCAGCTTAAAAAGATACGCGGCGTGTG 648
655 GCAGTATGAGATGATTAACCTGCGCAACGAGCTAGTATGATGAACAGGTTACTC 714
649 CCGGCAAGATCTGAATCTTCAATCCTATGATTTACAGACGAAACAAACATGCTGACAT 708
715 ACAGACCAATCCCAAGGCTTTCTTAACCTGCAATAGTATGACCAAGATCTGATTTG 774
709 GTTGTGTTGAGTCACATCGGGGCAAGATATGACAAATGATCAACGCGCAGCGC 768
775 GTTATGCGCTATATTCATGCTGAATAGATATATTTGCTCAACCTGATGCAACACAC 834
769 CACCTTGAAGAGCATGTCTGATGCGGAGCTGACATCATTCGTGCGCATCATCCGAC 828
835 GAGCTGCTGAAGCTTAATGATGCTGAGTATGTTGTTATTTGCTCCACACACAC 894
829 GTCTTGAACCGATGAAATGATATTAACGAAACGCTCATTTTCTACAGCTTCGCACTTT 888
895 TCCCTCTGCTGTTGAAAAATCCAAAGATGGATTTATTTTATATGCTTGGGCAATTTT 954
889 GTCTTGAACAGGCTGAGAGAAACAGAGACATGACATGCTGATGATCACTTAAG 948
955 ATTTTGAACCAAGGATGCAAAAGTTCAACGATTCGTTATTTTGAACATGATATCAAT 1014

```

```

OY 949 AAAATGAGACAGCGCGCTTGAAGTACACCGATGATATTCATGAACGACACTGCA 1008
DB 1015 AGTCCAGACGAGTTTCGTTTACCTTAAGGCCAATGAATAAGAGCTATTCACAAA 1074
OY 1009 CC 1010
DB 1075 CC 1076

RESULT 10
ADB06119
ID ADB06119 standard; DNA; 1212 BP.
AC ADB06119;
XX 20-NOV-2003 (first entry)
DE
XX Allotiococcus ostitis antigenic protein encoding DNA SEQ ID NO:59.
DE
XX Allotiococcus ostitidis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection; gene; de.
XX Allotiococcus ostitis.
OS
XX MO2003048304-A2.
XX
XX 12-JUN-2003.
XX
XX 25-NOV-2002; 2002WO-US036123.
XX
XX 29-NOV-2001; 2001US-0333777P.
XX 18-NOV-2002; 2002US-0426742P.
XX (AMBP) WYETH HOLDINGS CORP.
XX Fletcher LD, McMichael JC, Russell DF, Zagursky RJ;
XX WPI, 2003-505284/47.
XX P-PSDB; ADB06120.
XX
XX New Allotiococcus ostitidis polynucleotides and polypeptides, useful for
XX treating and diagnosing diseases, drug screening assays and monitoring of
XX effects during drug clinical trials.
XX
XX Claim 7, SEQ ID NO 59, 1019pp, English.
XX
XX The present invention describes an isolated polynucleotide (1) of
XX Allotiococcus ostitidis genomic DNA, which encodes an antigenic protein.
XX Allotiococcus ostitidis is a Gram-positive bacterium. Also described: (1)
XX an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
XX expression vector comprising the novel isolated polynucleotide (1), its
XX complement, degenerate variant or fragment; (3) a genetically engineered
XX host cell, transfected, transformed or infected with the vector of (2);
XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
XX composition comprising the polypeptide, its complement, biological
XX equivalent or fragment, or the polynucleotide that is comprised in the
XX expression vector; (6) a pharmaceutical composition comprising the
XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array
XX of the polypeptides of (1), their biological equivalent or fragment; (8)
XX immunising against Allotiococcus ostitidis by administering to a host the
XX immunogenic composition; (9) detecting and/or identifying Allotiococcus
XX ostitidis in the biological sample; (10) a kit comprising a container
XX containing the novel polynucleotide, its degenerate variant or fragment,
XX or the antibody of (4); and (11) producing a polypeptide by culturing the
XX genetically engineered host cell under conditions suitable to produce the
XX polypeptide from the culture. (1) can be used in gene therapy. The
XX polynucleotides, polypeptides, antibodies and compositions of the present
XX invention can be used for treating and diagnosing diseases, drug
XX screening assays and monitoring of effects during drug clinical trials.
XX The polynucleotides are useful for expressing and detecting Allotiococcus
XX ostitidis. The present sequence encodes an Allotiococcus ostitidis antigen
XX protein from the present invention.

```

SQ Sequence 1212 BP; 353 A; 248 C; 259 G; 352 T; 0 U; 0 Other;
 Query Match 8.5%; Score 96.4; DB 9; Length 1212;
 Best Local Similarity 46.7%; Pred. No. 7.8e-17;
 Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

QY	184	TCATTGTAGGCGATATATATGATGAGACGCTATGTTGAAAAAGTAAACGAGCAAAAAAGG	243
DB	238	TCATTGTGGGCGATATATCCTTGAAGTCCTATATATGGGAAATCGGGAACGAGATGCT	297
QY	244	GCAGACGATATTTTCAATATATGTAACCGATCTTTAGAGCTCGGATTAATGACGAGA	303
DB	298	TACGAGAAATTTTACATATGTTAAACCATTTTGAACGATGAGACTTGGTCTTGCC	357
QY	304	AACCTTGAACCCGGTAACCTATCAAAAGAAATTATAA-----ACAAGCA	348
DB	358	AATCTGATATCGCGTACCTATGACATGAACTTACGTTAAGCCGGCTTCAAGTTTA	417
QY	349	GATAAAGATTCATCTGACAGACGAAATGAAGATCAAGTAAAGTCTTGAAGGATATGAT	408
DB	418	AGTGAGGATTTATTTATGATTTCTTCATGACTGACGCAAAAGCATGCAAGAGCTGGC	477
QY	409	TTACAGGTTCTCAACAGCCCAACACCAACGATGATTAACGCGCTTCAGGCGATGAA	468
DB	478	ATTGACCTAGTTTCCATGCGCAACACCATACCGGCGACATGGGCAACAGGATGCTT	537
QY	469	GATACGCTTGAGATTTGCGAAGCAAAACCTGATATCGTTGAGCGGAGTACAGCTTA	528
DB	538	GACGCGATGGAATTTTAAAGATGTAAAGATGATTAATTTGGAATGGGCGCATGATCGA	597
QY	529	AGTGTGCGAAAAAGAAATTTCTGACAGAAAGTCAACGGGGTAAAGATTGCAACGCTT	588
DB	598	GCAGAGC---TGCGCAAGCTTATCAATTTGTGGGATTAATTTGACACCTCCATTTT	654
QY	589	GCGTTTACGATGTCTCCGGGAAGGTTTCCGGGCTTAAAAAGATACCCCGGCGCTGTG	648
DB	655	GCAGTTACCGATGTGATTTAACTGGCCACACGCTAGTGTATGAACACAGGTGTACTC	714
QY	649	CCCCGAGTCTGAAATCTTCATCCCTATGATTTGATGAAGGAAAAACATGCTGACATT	708
DB	715	ACGACCAATCTCCAAACCTTTCTTAACCTGGCCAAATAGCTATGACCAAGAAATCTGATTG	774
QY	709	GTGTTGTGCACTGCACTGCGGCGCAAGATATGACATGATTCMAACGACCGCCAGGC	768
DB	775	GTTATCGGCTATATCTGATGCTGGAATAGATATATTCCTCAACCTGATGCAACACAG	834
QY	769	CAGCTTGCAAGACCATGTCTGATGCGGAGCTGACATCATGCTGCGCCATCATCCGAC	828
DB	835	GAGCTGGCTGAAAGCTTAATGATCTGTGTCAGATATTGTTATTTGCTCCACACCCAC	894
QY	829	GCTTTAGAACCGATGAGTATATTAACGGAACCGTCAATTTTCTACAGCTCGGCAACTT	888
DB	895	TCCCTCTGCTGCTGGAATAAATACCAAGATGGGATTTATTTTATGAGCTTGGCAATTT	954
QY	889	GCTTTGACCAAGCTGACGAGCAAGAACAGACAGTGTGCTGATATCACTGGAAG	948
DB	955	ATTTTTCACAAAGGATGCAAAAGTTCAACCATTCCTGATTTTGAATGATGATCAAT	1014
QY	949	AAAAATGACAGCGCGCTTGAAGTGAACCCGATGATATCATGAAGCAACCTGCA	1008
DB	1015	AGTCACAGACAGTTGTTTACCTTAAGGCCAATGAAATGAAGCTGTATTCCTCAAA	1074
QY	1009	CC 1010	
DB	1075	CC 1076	

RESULT 11
 ADB06117
 ID ADB06117 standard; DNA; 1212 BP.
 AC ADB06117;
 ,XX

DT 20-NOV-2003 (first entry)
 DE Allotiococcus otilidis antigenic protein encoding DNA SEQ ID NO:57.
 XX
 XX Allotiococcus otilidis; antigenic protein; immunogenic; immunisation;
 KM gene therapy; Gram-positive bacterium; infection; gene; ds.
 XX
 XX Allotiococcus otilidis.
 PN WO2003048304-A2.
 XX
 PD 12-JUN-2003.
 XX
 PF 25-NOV-2002; 2002WO-US036123.
 PF
 PR 29-NOV-2001; 2001US-0333777P.
 PR 18-NOV-2002; 2002US-0426742P.
 XX
 PA (AMHP) WYETH HOLDINGS CORP.
 PI Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ,
 XX
 DR WPI; 2003-505284/47.
 DR P-PSDB; ADB06118.
 XX
 PT New Allotiococcus otilidis polynucleotides and polypeptides, useful for
 PT treating and diagnosing diseases, drug screening assays and monitoring of
 PT effects during drug clinical trials.
 XX
 PS Claim 7; SEQ ID NO 57; 1019pp; English.

The present invention describes an isolated polynucleotide (1) of
 CC Allotiococcus otilidis genomic DNA, which encodes an antigenic protein.
 CC Allotiococcus otilidis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
 CC expression vector comprising the novel isolated polynucleotide (1), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against Allotiococcus otilidis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying Allotiococcus
 CC otilidis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (1) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during clinical trials.
 CC The polynucleotides are useful for expressing and detecting Allotiococcus
 CC otilidis. The present sequence encodes an Allotiococcus otilidis antigen
 CC protein from the present invention.

SQ Sequence 1212 BP; 353 A; 248 C; 259 G; 352 T; 0 U; 0 Other;
 Query Match 8.5%; Score 96.4; DB 9; Length 1212;
 Best Local Similarity 46.7%; Pred. No. 7.8e-17;
 Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

QY	184	TCATTGTAGGCGATATATATGATGAGACGCTATGTTGAAAAAGTAAACGAGCAAAAAAGG	243
DB	238	TCATTGTGGGCGATATATCCTTGAAGTCCTATATTCGGAATCGGTGAACGAGATGT	297
QY	244	GCAGACGATATTTTCAATATATGTAACCGATCTTTAGAGCTCGGATTAATGACGAGA	303
DB	298	TACGAGAAATTTTACATATGTTAAACCATTTTGAACGATGAGACTTGGTCTTGCC	357
QY	304	AACCTTGAACCCGGTAACCTATCAAAAGAAATTATAA-----ACAAGCA	348

```

DB 358 AATCTTGAATTCGCGCTACCTATGCAATGCAACTTACGTTACCGGCTTCAAGTTTA 417
OY 349 GATTAAGAGATTCATCTGACAGCAATTAAGAAATGAGAAAGTCTTGAAGATATGAAAT 408
DB 418 AGTGAAGGATTTATTTGATTTCTTCCATGACTAGCAAAAGCATGCAAGAACTGGAC 477
OY 409 TTCACGGTTCTCAACAGGCGCAACCAACGATGATTAACGGGCTTCAAGGCAATGAAA 468
DB 478 ATTGAACCTAGTTTCCATGCGCAACCACTACCGGCAATGGGCAAGAGGATGTT 537
OY 469 GATAGGCTTGGAGATTTGGCAAGCAAACTTATGCTTGAAGCGGATACAGCTTA 528
DB 538 GACGGCAATGGAATTTTAAAGATGTAATGATGATTAATGGAATGGCCATGATGCA 597
OY 529 AGTATGCGAAAGAAATTTCTGACCAAGAAAGTCAACGGGATACGATTCACACGCTT 588
DB 598 GCGAAGAC---TGGCCAGCTTATCAATTTGTTGGAAATTAATTTGACTACCTCAATTTT 654
OY 589 GCGTTTACCGATGTCTCCGGGAAAGGTTTCCGCGCTAAAGAAATACCGCGGCTGCTG 648
DB 655 GCGATTAGGATGATGATTAACCTGGCCAAAGCAGTATGATGATGATGATGATGATGATC 714
OY 649 CCGCGAGATTCGAAATCTTCACTCCCTATGATTTGCAAGCAAGCAAAACATGCTGACAT 708
DB 715 AGACCAACTCCCAAGCTTCTTAACTGGCCCAATAGCTATGACCAAGAAATCTGATTTG 774
OY 709 GTTGTGTGACATGACACTGGGCGCAAGATGACATGATCAACAGCAAGCGCCAGCC 768
DB 775 GTTATGCGCTATATCTCATCTGGAATAGATTAATTTGCTCAACCTGATCAACAGCAAG 834
OY 769 CAGCTTGCAAGAGCCATGCTGATGCGGAGACTGACATCATGCTCGGCATCATCCGAC 828
DB 835 GAGCTGCTGAAAGCTTAATGATGCTGGTGACGATGATGATGATGATGATGATGATGAT 894
OY 829 GCTTATGAAACCATTAATGATTAATGACGAAACCGTATTTTCAAGCGCTCGGCACTT 888
DB 895 TCCCTCTCTCTGTTTAAATAATCAAGATGGATTAATTTTATGCTTGGCAATTTT 954
OY 889 GTCTTGACCAAGGCTGACGAGAACAGACAGTCACTGGTATGATCACTGAAG 948
DB 955 ATTTTGACCAAGGCTGACGAAATTTCAACGATTTCCGTTATTTTAAATGATGATCAAT 1014
OY 949 AAAAATGAAACAGCGCTTGTGAAGTACACCGATGATATCCATGAAGCAACCTGCA 1008
DB 1015 AGCCGACACGAGTTGTTTACCTTAAGCCCAATGAATAAGCTGATTTCCCAAA 1074
OY 1009 CC 1010
DB 1075 CC 1076

```

```

RESULT 12
ADBI2064_00/c
WP Sequence split into 18 fragments LOCUS ADBI2064 Accession Adbi2064
WP Fragment Name Begin End
WP ADBI2064_00 1 110000
WP ADBI2064_01 100001 210000
WP ADBI2064_02 200001 310000
WP ADBI2064_03 300001 410000
WP ADBI2064_04 400001 510000
WP ADBI2064_05 500001 610000
WP ADBI2064_06 600001 710000
WP ADBI2064_07 700001 810000
WP ADBI2064_08 800001 910000
WP ADBI2064_09 900001 1010000
WP ADBI2064_10 1000001 1110000
WP ADBI2064_11 1100001 1210000
WP ADBI2064_12 1200001 1310000
WP ADBI2064_13 1300001 1410000
WP ADBI2064_14 1400001 1510000
WP ADBI2064_15 1500001 1610000
WP ADBI2064_16 1600001 1710000

```

```

WP ADBI2064_17 1700001 1754382
ID ADBI2064 standard; DNA; 1754382 BP.
XX
AC ADBI2064;
XX
DT 20-NOV-2003 (first entry)
XX
DE Alloicoccus oclidis entire genome sequence SEQ ID NO:6651.
XX
KM Alloicoccus oclidis; antigenic protein; immunogenic; immunisation;
XX gene therapy; Gram-positive bacterium; infection; gene; ds.
XX
OS Alloicoccus oclidis.
XX
PN WO2003048304-A2.
XX
PD 12-JUN-2003.
XX
PF 25-NOV-2002; 2002WO-US036123.
XX
PR 29-NOV-2001; 2001US-0333777P.
XX 18-NOV-2002; 2002US-0426742P.
XX
PA (AMHP ) WYETH HOLDINGS CORP.
XX
PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ,
XX
DR WPI; 2003-505284/47.
XX
XX
XX New Alloicoccus oclidis polynucleotides and polypeptides, useful for
XX treating and diagnosing diseases, drug screening assays and monitoring of
XX effects during drug clinical trials.
XX
XX Example 3; SEQ ID NO 6651; 1019pp; English.
XX
XX
XX The present invention describes an isolated polynucleotide (1) of
XX Alloicoccus oclidis genomic DNA, which encodes an antigenic protein.
XX Alloicoccus oclidis is a Gram-positive bacterium. Also described: (1)
XX an isolated polypeptide, that is encoded by the polynucleotide (1); (2) an
XX expression vector comprising the novel isolated polynucleotide (1), its
XX complement, degenerate variant or fragment; (3) a genetically engineered
XX host cell, transfected, transformed or infected with the vector of (2);
XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
XX composition comprising the polypeptide, its complement, biological
XX equivalent or fragment, or the polynucleotide that is comprised in the
XX expression vector; (6) a pharmaceutical composition comprising the
XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array
XX of the polypeptides of (1), their biological equivalent or fragment; (8)
XX immunising against Alloicoccus oclidis by administering to a host the
XX immunogenic composition; (9) detecting and/or identifying Alloicoccus
XX oclidis in the biological sample; (10) a kit comprising a container
XX containing the novel polynucleotide, its degenerate variant or fragment,
XX or the antibody of (4); and (11) producing a polypeptide by culturing the
XX genetically engineered host cell under conditions suitable to produce the
XX polypeptide from the culture. (1) can be used in gene therapy. The
XX polynucleotides, polypeptides, antibodies and compositions of the present
XX invention can be used for treating and diagnosing diseases, drug
XX screening assays and monitoring of effects during drug clinical trials.
XX The polynucleotides are useful for expressing and detecting Alloicoccus
XX oclidis. The present sequence represents the entire genome of
XX Alloicoccus oclidis, which is given in the exemplification of the
XX present invention.
XX
XX Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 00Other;
XX
XX
XX Query Match 8.5%; Score 96.4; DB 9; Length 110000;
XX Best Local Similarity 46.7%; Pred No. 6 2e-16;
XX Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;
OY 184 TCATTGTGAGCATATTAATGATGAGCAGCTATGTTGAAAAAGTAAAGGAGCAAAAGGG 243
DB 24408 TCATTGTGAGCATATTAATGATGAGCAGCTATGTTGAAAAAGTAAAGGAGCAAAAGG 24349

```


DB 46698 GTTATTTGTTATATGCTACCTGGGCGCTCCGATATCTGTGAACCTCCAAACCAACCAACT 46639
QY 769 CAGCTTGCAGAGAGCCATGCTGATGCGGAGAGCTGACATCATGCTGGCCATCATCCGAC 828
DB 46638 CAAATTTGACATGCTATTTTAAATGCTGCTGATATATATATGAGAGCTCATCCCCAC 46579
QY 829 GTCTTAGAACCCATTTGAAGTATATACGAGACCCGTCATTTTCTACAGCCTCGGCAACTT 888
DB 46578 CCGCTGGAAGATGCGAAAAATATAAATATATATGTTATACAGATATGCGCATTTT 46519
QY 889 GTCTTTGAC 897
DB 46518 GCTTTTGGC 46510

RESULT 15

AB068713 standard; DNA; 885 BP.

AB068713;

29-AUG-2003 (revised)

29-AUG-2002 (first entry)

Listeria monocytogenes 4b contig DNA sequence #1479.

Antibacterial; Listeria; food contamination; mutational analysis; infection; ds.

Listeria monocytogenes ATCC 19115.

W0200228891-A2.

11-APR-2002.

04-OCT-2001; 2001MO-FR03061.

04-OCT-2000; 2000FR-00012697.

(INSP) INST PASTEUR.

(CNRS) CNRS CENT NAT RECH SCI.

Kunst F, Glaser P;

WPI; 2002-332479/37.

New genomic sequences from Listeria species, useful for detection, treatment and prevention of infection, also related polypeptides, antibodies and modulators.

Claim 14; SEQ ID NO 1526; 180bp; French.

CC The present invention relates to nucleic acid sequences (AB067188-
CC AB071212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (updated
CC on 29-AUG-2003 to standardise OS field)

SQ Sequence 885 BP; 275 A; 183 C; 161 G; 266 T; 0 U; 0 Other;

Query March 6.5%; Score 74; DB 6; Length 885;

Best Local Similarity 54.4%; Pred. No. 1.9e-10;

Matches 149; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

DB 40 GCAGACAGCCATACCCAGCCGATATCCCATGCTGATATTTGAAAGATTTGATGAGAGCTG 99
QY 682 TCAGAGCGGAAAAAACAATGCTGACATTTGTTGTGAGTCACTGCGGCAAGATAT 741
DB 100 AAGACATACAAAAAAGAACTCTGCTGTTATCATATATACGACTGGGGGCTGGAATAT 159
QY 742 GACAATGATCCAAACGACCGCCAGCCAGCTTGCAGAGGCCATGTGATGCGGAGCT 801
DB 160 CGCGAAACACCAACGATTTATCAAAACCAATTTGGTCATGCGATTTAGATGCTGTCCT 219
QY 802 GACATCATGTCGGGCATCATCCGACGCTTTAGAACGATTGAAGTATATACGAACC 861
DB 220 GATATATTTATGGGCTCTCACCTCATGCGCTAGAAAGCGTTGAAAAATATATAAAGATAA 279
QY 862 GTCAATTTCTACAGCTCGGCACTTGTCTTTG 895
DB 280 TACATGTTTATAGCATGCGGCACTTGTCTTTG 313

Search completed: February 27, 2006, 07:05:22
Job time : 723.889 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model.

Run on: February 27, 2006, 06:34:28 ; Search time 1248.33 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-3
Perfect score: 1140
Sequence: 1 acgaaagaactgagctt.....gtgcaactaactaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications NA Main:
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1140	100.0	1140	US-10-484-605-3	Sequence 3, Appl1
2	1140	100.0	1140	US-10-789-164-3	Sequence 3, Appl1
3	1140	100.0	6536	US-10-484-605-6	Sequence 6, Appl1
4	332.4	29.2	636	US-09-974-300-1745	Sequence 1745, Ap
5	225.6	19.8	488	US-09-974-300-1777	Sequence 1777, Ap
6	96.4	8.5	1212	US-10-501-282-55	Sequence 55, Appl
7	96.4	8.5	1212	US-10-501-282-57	Sequence 57, Appl
8	96.4	8.5	1212	US-10-501-282-59	Sequence 59, Appl
9	96.4	8.5	1754382	US-10-501-282-6651	Sequence 6651, Ap
10	77.8	6.8	1163020	US-10-398-221-10	Sequence 10, Appl
11	77.8	6.8	3011208	US-10-398-221-2058	Sequence 2058, Ap
12	74	6.5	885	US-10-398-221-1526	Sequence 1526, Ap
13	74	6.5	2282	US-10-398-221-3371	Sequence 3371, Ap
14	54.2	4.8	1071	US-10-470-0488-514	Sequence 514, Ap
15	54.2	4.8	1101	US-10-724-972A-1055	Sequence 1055, Ap
16	47.6	4.3	1293	US-10-474-792-53	Sequence 86, Appl
17	47.6	4.2	1170	US-10-724-972A-866	Sequence 866, Appl
18	47.2	4.1	536165	US-09-939-964-1	Sequence 1, Appl1
19	41.8	3.7	2225	US-11-097-143-34615	Sequence 34615, A
20	41.8	3.7	3256	US-10-104-047-474	Sequence 474, Ap
21	39.6	3.5	1467	US-10-501-282-2261	Sequence 2261, Ap
22	39.6	3.5	1467	US-10-501-282-2263	Sequence 2263, Ap
23	39.6	3.5	1754382	US-10-501-282-6651	Sequence 6651, Ap

C 24	39	3.4	620	4	US-09-925-065A-700467	Sequence 700467,
C 25	39	3.4	620	4	US-09-925-065A-700468	Sequence 700468,
C 26	39	3.4	1019	3	US-09-981-566A-96	Sequence 96, Appl
C 27	39	3.4	1205	7	US-10-424-599-107806	Sequence 107806,
C 28	39	3.4	1348	6	US-10-017-161-1059	Sequence 1059, Ap
C 29	39	3.4	1348	6	US-10-292-798-901	Sequence 901, Appl
C 30	38.6	3.4	1921	3	US-09-764-868-161	Sequence 161, Appl
C 31	38.6	3.4	2789	7	US-10-161-493-27	Sequence 27, Appl
C 32	38.6	3.4	2812	6	US-10-161-493-27	Sequence 6177, Ap
C 33	38.4	3.4	831	6	US-10-033-585-6177	Sequence 935, Appl
C 34	38	3.3	437	9	US-10-915-740A-935	Sequence 5305, A
C 35	38	3.3	549	4	US-09-925-065A-55305	Sequence 7, Appl1
C 36	38	3.3	56609	9	US-10-915-740A-7	Sequence 1068, Ap
C 37	38	3.3	2242716	9	US-10-915-740A-1068	Sequence 634, Appl
C 38	37.4	3.3	606	9	US-10-617-320-634	Sequence 6837, Ap
C 39	37.2	3.3	510	6	US-10-029-386-6837	Sequence 16363, A
C 40	37.2	3.3	544	7	US-10-021-323-16363	Sequence 14889, A
C 41	37.2	3.3	590	7	US-10-021-323-14869	Sequence 5675, Ap
C 42	37.2	3.3	814	8	US-10-767-795-5675	Sequence 174562,
C 43	37.2	3.3	2622	8	US-10-425-115-174562	Sequence 1036, Ap
C 44	37.2	3.3	3710	6	US-10-294-025-1036	Sequence 83576, A
C 45	37	3.2	1133	7	US-10-424-599-83576	

ALIGNMENTS

RESULT 1
US-10-484-605-3
; Sequence 3, Application US/10484605
; Publication No. US20040253704A1
; GENERAL INFORMATION:
; APPLICANT: Sung, Moon-Hee
; APPLICANT: Hong, Seung-Pyo
; APPLICANT: Lee, Jong-Su
; APPLICANT: Jung, Chang-Min
; APPLICANT: Kim, Chul-Joong
; APPLICANT: Soda, Kenji
; APPLICANT: Ashikichi, Makoto
; TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING pgsBcA, THE GENE CODING
; TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
; FILE REFERENCE: 4240-101
; CURRENT APPLICATION NUMBER: US/10/484,605
; PRIOR FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: PCT/KR02/01522
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: KR 2001-48373
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-484-605-3
Query Match 100.0%; Score 1140; DB 8; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTGCTAAAGCTGACAAACGCAAAAAAG 60
DB 1 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTGCTAAAGCTGACAAACGCAAAAAAG 60
QY 61 AAAACCAATGACGATTTATTGCAATTCGATGCTTTTGTCTTATGTTGCTTTC 120
DB 61 AAAACCAATGACGATTTATTGCAATTCGATGCTTTTGTCTTATGTTGCTTTC 120
QY 121 ATGTGGCGGGAAGCGAAACGCGGAAGCTCAAAAGCTATTCTGACGACGATCTCA 180
DB 121 ATGTGGCGGGAAGCGAAACGCGGAAGCTCAAAAGCTATTCTGACGACGATCTCA 180

OY	181	GCCTCATTTGAGCGAATAATATGATGGGACGCTATGTTGAAAAGTAACGACACAAAA	240
Db	181	GCTCATTTTGAAGCGAATTTATGATGGGACGCTATGTTGAAAAGTAACGACACAAAA	240
OY	241	GGGCGACAGCATATTTTCAATATATGTGAACCGATCTTTAGAGCTCGGATTAATGACGA	300
Db	241	GGGCGACAGCATATTTTCAATATATGTGAACCGATCTTTAGAGCTCGGATTAATGACGA	300
OY	301	GGAAACTTTGAAAACCCGCTAACCTATCAAAAGAAATTAACAAGCAGATTAAGAGATT	360
Db	301	GGAAACTTTGAAAACCCGCTAACCTATCAAAAGAAATTAACAAGCAGATTAAGAGATT	360
OY	361	CATCTGCAGAGAAATTAAGGAATCAAGTAAAGTCTTTGAAAGATTAATTTACGGTCTC	420
Db	361	CATCTGCAGAGAAATTAAGGAATCAAGTAAAGTCTTTGAAAGATTAATTTACGGTCTC	420
OY	421	AACAGCGCCACCAACACAGCAATGGAATTCAGCGCTTCAGGCGATGAAGATACGCTTGA	480
Db	421	AACAGCGCCACCAACACAGCAATGGAATTCAGCGCTTCAGGCGATGAAGATACGCTTGA	480
OY	481	GAATTTGGGAAGCAAAACCTTGATATCGTTGGAGCGGGATACAGCTTAAGTGATCCGAA	540
Db	481	GAATTTGGGAAGCAAAACCTTGATATCGTTGGAGCGGGATACAGCTTAAGTGATCCGAA	540
OY	541	AAGAAATTTTGSTGCCAGAAAGTCACGCGGTAAAGATGGAACGCTGGCTTTAACGAT	600
Db	541	AAGAAATTTTGSTGCCAGAAAGTCACGCGGTAAAGATGGAACGCTGGCTTTAACGAT	600
OY	601	GTGTCCGGGAAAAGTTTTCGCGCTAAAAAGATAAGCCGGGCGTGCCTGCCAGATCT	660
Db	601	GTGTCCGGGAAAAGTTTTCGCGCTAAAAAGATAAGCCGGGCGTGCCTGCCAGATCT	660
OY	661	GAATCTTCAATCCCTATATGATTTCAAGACGAAAAACAATGCTGACATTTGTTGTGCAG	720
Db	661	GAATCTTCAATCCCTATATGATTTCAAGACGAAAAACAATGCTGACATTTGTTGTGCAG	720
OY	721	TCACACTGGGGCCAAAGATATGACAAATGATCCAAACGACCGCCAGCGCAGCTTGCAAGA	780
Db	721	TCACACTGGGGCCAAAGATATGACAAATGATCCAAACGACCGCCAGCGCAGCTTGCAAGA	780
OY	781	GCCATGTCGATGCGGGAGCTGACATCATCTGCGGCATCATCCGCAAGCTTTAGAACCG	840
Db	781	GCCATGTCGATGCGGGAGCTGACATCATCTGCGGCATCATCCGCAAGCTTTAGAACCG	840
OY	841	ATTGAAGTATTTAACGGAACCGTCATTTTTCACAGCTTCGGCAACTTTGCTTTGACCCA	900
Db	841	ATTGAAGTATTTAACGGAACCGTCATTTTTCACAGCTTCGGCAACTTTGCTTTGACCCA	900
OY	901	GGCTGGAGAGAAACAAGACAGTGCATGCTGTCAGTATCACTGAAAGAAAATGGAACA	960
Db	901	GGCTGGAGAGAAACAAGACAGTGCATGCTGTCAGTATCACTGAAAGAAAATGGAACA	960
OY	961	GGCGGCTTTGAAGTGACACCGATCGATATCAAGAGGACACCTGCACCTGTGAAAAAA	1020
Db	961	GGCGGCTTTGAAGTGACACCGATCGATATCAAGAGGACACCTGCACCTGTGAAAAAA	1020
OY	1021	GACAGCTTTAAACAGAAAACCATTAATTCGCAACGACGAAAGACTTAATTCGCTTGG	1080
Db	1021	GACAGCTTTAAACAGAAAACCATTAATTCGCAACGACGAAAGACTTAATTCGCTTGG	1080
OY	1081	AAAGTAGAAGACGGAACCTGACGTTGATATTTGATTCATACGCAAACTAAATCTTAA	1140
Db	1081	AAAGTAGAAGACGGAACCTGACGTTGATATTTGATTCATACGCAAACTAAATCTTAA	1140

RESULT 2
US-10-789-164-3
; Sequence 3, Application US/10789164
; Publication No. US2005019120A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENETICALLY MODIFIED BACTERIAL CELL
; TITLE OF INVENTION: ENCODING POLY-GAMMA-GLUTAMATE SYNTHETASE

```

: FILE REFERENCE: PI574
: CURRENT APPLICATION NUMBER: US/10/789,164
: CURRENT FILING DATE: 2004-02-27
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: Koparentin 1.71
: SEQ ID NO: 3
: LENGTH: 1140
: TYPE: DNA
: ORGANISM: Bacillus subtilis
US-10-789-164-3

```

Query Match	100.0%	Score 1140	DB 9	Length 1140
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1140	Conservative 0	Mismatches 0	Indels 0	Gaps 0

OY	1	ATGAAAAAGAACTGACCTTTCATGAAACCTGTAAAGCTGACAAACAGCAAAAAAG	60
Db	1	ATGAAAAAGAACTGACCTTTCATGAAACCTGTAAAGCTGACAAACAGCAAAAAAG	60
OY	61	AAAACCAATAGACAGTATTATATGTCATTCGATCGTTTTGTCTTATGTTGCTTTC	120
Db	61	AAAACCAATAGACAGTATTATATGTCATTCGATCGTTTTGTCTTATGTTGCTTTC	120
OY	121	ATGTGGGCGGGAAAGCGGGAAACGCCCAAGTCAAAACGTATTTCTGACGATCTTCA	180
Db	121	ATGTGGGCGGGAAAGCGGGAAACGCCCAAGTCAAAACGTATTTCTGACGATCTTCA	180
OY	181	GCCTCATTTTGTAGGCGATATTATATGATGGGACGCTATGTTGAAAAAGTAAAGSAGCAAAAA	240
Db	181	GCCTCATTTTGTAGGCGATATTATATGATGGGACGCTATGTTGAAAAAGTAAAGSAGCAAAAA	240
OY	241	GGGGCAACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTTCGATTTATGACGA	300
Db	241	GGGGCAACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTTCGATTTATGACGA	300
OY	301	GGAAACTTTGAAAAACCGGTAAOCTATCAAAAGAAATTATAACAAGAGATTAAGAAGATT	360
Db	301	GGAAACTTTGAAAAACCGGTAAOCTATCAAAAGAAATTATAACAAGAGATTAAGAAGATT	360
OY	361	CATCTGCAGACGAATTAAGAAATCAGTAAAGTCTTGAAGGATATGATTTTCAAGGTTCTC	420
Db	361	CATCTGCAGACGAATTAAGAAATCAGTAAAGTCTTGAAGGATATGATTTTCAAGGTTCTC	420
OY	421	AACAGCGCCAAACAACGCAATGATTAACGGCTTCAGGGCATGAAGATACGCTTTGGA	480
Db	421	AACAGCGCCAAACAACGCAATGATTAACGGCTTCAGGGCATGAAGATACGCTTTGGA	480
OY	481	GAATTTGCGAAGCAAAACCTTGATATGCTTGGAGCGGGATATACGCTTAAGATAGCGAAA	540
Db	481	GAATTTGCGAAGCAAAACCTTGATATGCTTGGAGCGGGATATACGCTTAAGATAGCGAAA	540
OY	541	AAAGAAATTTTGTACAGAAAGTCAACGGGGTAAACGATTGCAACGTTTGCTTAACGAT	600
Db	541	AAAGAAATTTTGTACAGAAAGTCAACGGGGTAAACGATTGCAACGTTTGCTTAACGAT	600
OY	601	GTGTCCGGGAAAGGTTTTCCGGGCTAAAAAGATACGCCGGCGTGTCTGCCCGCAGATCTT	660
Db	601	GTGTCCGGGAAAGGTTTTCCGGGCTAAAAAGATACGCCGGCGTGTCTGCCCGCAGATCTT	660
OY	661	GAATTTCTTCATCCTTATGATTTTCGAAGCGAAAAAACAATGCTGACATTTGTTGTGTGCAG	720
Db	661	GAATTTCTTCATCCTTATGATTTTCGAAGCGAAAAAACAATGCTGACATTTGTTGTGTGCAG	720
OY	721	TCACACTGGGGCCAAAGAGTATGACAATGATCCAAACAGCCGACGCCACGCTTGCAGAGA	780
Db	721	TCACACTGGGGCCAAAGAGTATGACAATGATCCAAACAGCCGACGCCACGCTTGCAGAGA	780
OY	781	GCCATGTCGTATGCGGGAGCTGACATCATGCTGCGGCATCATTCGCGACAGTCTTGAACCG	840
Db	781	GCCATGTCGTATGCGGGAGCTGACATCATGCTGCGGCATCATTCGCGACAGTCTTGAACCG	840
OY	841	ATTGAAGTATTAACGAAACCGGTATTTTCTACAGCCTCGGCACATTTTGTCTTGAACCA	900
Db	841	ATTGAAGTATTAACGAAACCGGTATTTTCTACAGCCTCGGCACATTTTGTCTTGAACCA	900

Db 841 ATTGAAGTATATACGGAACCGTCATTTTCTACAGCCTCCGCACTTTGTCTTGACCA 900
Qy 901 GGCTGACGAGAAACAAGACAGTGCCTGCTGATACCTGAGAAAATGGAACA 960
Db 901 GGCTGACGAGAAACAAGACAGTGCCTGCTGATACCTGAGAAAATGGAACA 960
Qy 961 GGCGCTTTGAAGTACACCGATCCATATCCATGAAAGCAGACCTGCACCTGTGAAAAA 1020
Db 961 GGCGCTTTGAAGTACACCGATCCATATCCATGAAAGCAGACCTGCACCTGTGAAAAA 1020
Qy 1021 GACAGCCTTAAACAGAAAACATTAATGCGAAGCTGACGAAAGACTTAATTCCTTGG 1080
Db 1021 GACAGCCTTAAACAGAAAACATTAATGCGAAGCTGACGAAAGACTTAATTCCTTGG 1080
Qy 1081 AAAGTAGAAGACGGAACCTGACGTTTGAATTTGATGATGACAACTAAATCTTAA 1140
Db 1081 AAAGTAGAAGACGGAACCTGACGTTTGAATTTGATGATGACAACTAAATCTTAA 1140

RESULT 3

US-10-484-605-6
Sequence 6, Application US/10484605
Publication No. US20040253704A1

GENERAL INFORMATION:
APPLICANT: Sung, Moon-Hee
APPLICANT: Hong, Seung-Pyo
APPLICANT: Lee, Jong-Su
APPLICANT: Jung, Chang-Min
APPLICANT: Kim, Chul-Joong
APPLICANT: Soda, Kenji
APPLICANT: Ashiuchi, Makoto
TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING p85BCA, THE GENE CODING
TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
TITLE OF INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
FILE REFERENCE: 4240-101
CURRENT APPLICATION NUMBER: US/10/484,605
PRIOR FILING DATE: 2004-01-20
PRIOR APPLICATION NUMBER: PCT/KR02/01522
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: KR 2001-48373
PRIOR FILING DATE: 2001-08-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 6536
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Construct
US-10-484-605-6

Query Match 100.0%; Score 1140; DB 8; Length 6536;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 60
Db 1891 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 1950
Qy 61 AAAACCAATAGACAGTATTTATTTCCATTCGATCCGTTTGTCTTATGTTGCTTTC 120
Db 1951 AAAACCAATAGACAGTATTTATTTCCATTCGATCCGTTTGTCTTATGTTGCTTTC 2010
Qy 121 ATGTGGCGGGGAAAGCGGAAAGCGGAAAGTCAAAAGTATCTGACGAGTACTTCA 180
Db 2011 ATGTGGCGGGGAAAGCGGAAAGCGGAAAGTCAAAAGTATCTGACGAGTACTTCA 2070
Qy 181 GCCTCATTTGTAGCGATATTTATGATGAGCGCTATGTTGAAAAAGTAAAGCAAGCAAAA 240
Db 2071 GCCTCATTTGTAGCGATATTTATGATGAGCGCTATGTTGAAAAAGTAAAGCAAGCAAAA 2130
Qy 241 GGGGACAGACGATTTTTCATATGTTGAACCGATCTTTAGAGCTTCGATATGATGCA 300

Db 2131 GGGGACAGACGATTTTTCATATGTTGAACCGATCTTTAGAGCTTCGATATGATGCA 2190
Qy 301 GGAACCTTTGAAACCCGGTAAACCTATCAAAAGATTATAAACAAGCAGATTAAGAGATT 360
Db 2191 GGAACCTTTGAAACCCGGTAAACCTATCAAAAGATTATAAACAAGCAGATTAAGAGATT 2250
Qy 361 CATCTGACAGCAATTAAGCAATCAGTGAAGTCTTGAAGATGATGAATTTGACGGTCTC 420
Db 2251 CATCTGACAGCAATTAAGCAATCAGTGAAGTCTTGAAGATGATGAATTTGACGGTCTC 2310
Qy 421 AACAGCGCAACACACGCAATGATTAAGCGCTTCAGGCAATGAAAGATACGCTTGA 480
Db 2311 AACAGCGCAACACACGCAATGATTAAGCGCTTCAGGCAATGAAAGATACGCTTGA 2370
Qy 481 GAATTTGGAAAGCAAAACCTTGATATCTGTCGACCGGATACAGTTAAGTATGAGCGAAA 540
Db 2371 GAATTTGGAAAGCAAAACCTTGATATCTGTCGACCGGATACAGTTAAGTATGAGCGAAA 2430
Qy 541 AAGAAATTTGTAACAGAAAGTCAACGAGGTAAACGATTTGCAACGCTTGAACCGAT 600
Db 2431 AAGAAATTTGTAACAGAAAGTCAACGAGGTAAACGATTTGCAACGCTTGAACCGAT 2490
Qy 601 GTGTCCGGGAAAGGTTTTCGCGCTTAAAGAAATACCGCGCGTGTCTCCGCGAGATCT 660
Db 2491 GTGTCCGGGAAAGGTTTTCGCGCTTAAAGAAATACCGCGCGTGTCTCCGCGAGATCT 2550
Qy 661 GAATCTTTCATCTCTTATGATTTTCAAGACGGAAGAAACATGCTGACATTTGTTGTGCG 720
Db 2551 GAATCTTTCATCTCTTATGATTTTCAAGACGGAAGAAACATGCTGACATTTGTTGTGCG 2610
Qy 721 TCACACTGGGAGCAAGATGATGACAAATGATCCAAACGACCGGCAAGCTTTCAGAA 780
Db 2611 TCACACTGGGAGCAAGATGATGACAAATGATCCAAACGACCGGCAAGCTTTCAGAA 2670
Qy 781 GCCATGCTGATGCGGAGCTGACATCATGTCGCGCATCATGCGCAGCTTTAGAACCG 840
Db 2671 GCCATGCTGATGCGGAGCTGACATCATGTCGCGCATCATGCGCAGCTTTAGAACCG 2730
Qy 841 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCTTCGCAACTTTGTCTTTGACCA 900
Db 2731 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCTTCGCAACTTTGTCTTTGACCA 2790
Qy 901 GGCTGACGAGAAACAAGACAGTGCATGCTGCTGATACCTGAAAGAAATGGAACA 960
Db 2791 GGCTGACGAGAAACAAGACAGTGCATGCTGCTGATACCTGAAAGAAATGGAACA 2850
Qy 961 GGCGCTTTGAAGTACACCGATCCATATCCATGAAAGCAGACCTGCACCTGTGAAAAA 1020
Db 2851 GGCGCTTTGAAGTACACCGATCCATATCCATGAAAGCAGACCTGCACCTGTGAAAAA 2910
Qy 1021 GACAGCCTTAAACAGAAAACATTAATGCGAAGCTGACGAAAGACTTAATTCCTTGG 1080
Db 2911 GACAGCCTTAAACAGAAAACATTAATGCGAAGCTGACGAAAGACTTAATTCCTTGG 2970
Qy 1081 AAAGTAGAAGACGGAACCTGACGTTTGAATTTGATGATGACAACTAAATCTTAA 1140
Db 2971 AAAGTAGAAGACGGAACCTGACGTTTGAATTTGATGATGACAACTAAATCTTAA 3030

RESULT 4

US-09-974-300-1745
Sequence 1745, Application US/09974300
Patent No. US20020146721A1

GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06

PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1745
LENGTH: 696
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1745

Query Match 29.2%; Score 332.4; DB 3; Length 696;
Best Local Similarity 68.4%; Pred. No. 1.1e-85;
Matches 475; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

OY 444 GGATTAAGCGCCCTTCAGGCGATGAAGATACGCTTGGAATTTGGAGAGCAAACTTGA 503
DB 1 GGAATTAAGCGGAGAGCGGTTTGAAGATACGCTCAATTAATTTTCAATGAGATCTGA 60
OY 504 TATCGTTGAGCGGGATACAGCTTAAGTATGCGAAAAAGAAATTTGTACAGAAAGT 563
DB 61 GCTTGTGAGAGAGAAATTAATCTTGAAGACGCGAAACGATCTATCAGAAATGT 120
OY 564 CAACGGGGTAAAGATTGCAAGCTTGGCTTACCGATGTGTCGGGAAAGGTTTGGCGG 623
DB 121 GAACGGCGTAAATTTGCAAGCTGCTTTTACAGACGCTTACAGAAAGAACTTTACAGC 180
OY 624 TAAAGAAATACGCGGGCGTGTCCCGGAGATCTGAATCTTCAATCCCTATGATTC 683
DB 181 CAAAAGAACAGAGCGGAGTGTCCCGCTCAGTCCGAAATCTTATTCAGATATGTC 240
OY 684 AGAAGCGAAAAACATGCTGACATTTGTTGTGACATCACTGGGCGCAAGATATGA 743
DB 241 GGAAGCATCGAAAAAGCGGATCTTGTCTTGTCCATGTCAGTGGGCAAGATATGA 300
OY 744 CAATGATCAAAACGACCGCCAGCGGCTTGAAGCCATGTCTGATGCGGAGCTGA 803
DB 301 CAATGAACGGAACGACAGAGAAAGATCTGGCCAAAGCGATGTCAGATGTCGAGCAG 360
OY 804 CATATCGTCCGCGCATCATCCGCAAGTCTTGAACGATGATATTAACGAAACGT 863
DB 361 TGTCAATCGCGCGCTCATCTCCCATGTTCTCGAACCGATCGAAGGTATACGATACGT 420
OY 864 CATTTTCTACAGCCTCGGCACTTTGTCTTTGACCAAGGCTTGAAGAGAAACAGAG 923
DB 421 GATTTTCTACAGCCTCGGCACTTTGTATTTGATCAGGCGTGTGAAGAACGAGACG 480
OY 924 TGCATCTGTTCAATACCTGTAAGAAATGAAACGCGCTTTGAAATGACACCGAT 983
DB 481 CGCGCTTGTACAAATCAATTTAATGATGACGCAAGGCGCTTTGAGGTAAACCTCT 540
OY 984 CGATATCATGAAGGACACCTGCACCT---GTGAAAAAGACAGCGCTTAAACAGAAAC 1040
DB 541 CAACATTCGCGAAGACGCGGACGCTTTAGGCAAGCGACCTTTTAAACGAAAGC 600
OY 1041 CATTAATCGCGAAGTACGAAAGACTTAATTTCCCTTGAAGATGAGAGCGAAACT 1100
DB 601 GATCTTCGCTCAATGACAAAGAAACAACTCGACCTGGAAGAAAGAAAGAAATTT 660
OY 1101 GACGTTGATATGATCATAGTGAACAACTAAA 1134
DB 661 AACGTTGAAGTCATCATGCGGACAAAGCTGAA 694

RESULT 5
US-09-974-300-1777
Sequence 1777, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US

CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1777
LENGTH: 488
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1777

Query Match 19.8%; Score 225.6; DB 3; Length 488;
Best Local Similarity 67.8%; Pred. No. 1.1e-54;
Matches 331; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

OY 1 ATGAAAAAGAACTGACCTTTCATGAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
DB 1 ATGAAAAAGAACTGACCTTTCAGGAAAAAAGCTGTAAGTGAAGAGCAAGAAAAAG 60
OY 61 AAAACCAATAGACAGTATTTATTTCCATTCCTGATTCCTTTTGTCTTATGCTTC 120
DB 61 AAAACCAAGCAAGCGTCTTATGATGATGCGCTTATTTTGTCTTATGATGCTTC 120
OY 121 ATGAGGCGGGAAGGCGAAAGCG---CGAAGGTCAAAACGATTTCTGACGAGTATC 177
DB 121 ACTTGTGAGAGCGGCAAAATCTCTTCCGAAATGACAAAAAGAAAGATGCAAGCTT 180
OY 178 TCAGCTCATTTTGTAGCGATATATGATGAGGAGCTATGTTGAAAAAGTACGAGCA 237
DB 181 ACAGCTACTTTTGTGCGATATCATGATGGAAGAAAGCAAGTGAACAACTTG 240
OY 238 AAAGGCGACAGATATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATATGTA 297
DB 241 CACGTTTCGAAAGTGTCTTCAAAATGTGAAGCGGATCTTAAATGTGATGATTTATC 300
OY 298 GCAGAAACTTTGAAACCGGCTGACCTATCAAAAGATTAATAACAGCAGATTAAGAG 357
DB 301 ACAGAAACTTTGAAACCGGCTGATCAATGCAAAAGACATCAAGAGCGGAAAGAAC 360
OY 358 ATTCACTGACAGAAATTAAGAAATCAAGTAAGTCTTTGAAGATGATTTCAAGCTT 417
DB 361 ATCACTGCAAGCAATCAAGATCAAGTCAAAATTTGAAAAAGCTGAACTTCAAGCTA 420
OY 418 CTCACAGCGGCAACCAACGCAATGATTAAGGCTTCAAGGCGATGAAGATACGCTT 477
DB 421 CTGAATTTTGCACCAACCATGCGATGAGTACGAGGAGAGCGGTTTGAAGATACGCTC 480
OY 478 GGAGATT 485
DB 481 AATGATT 488

RESULT 6
US-10-501-282-55
Sequence 55, Application US/10501282
Publication No. US20050203280A1
GENERAL INFORMATION:
APPLICANT: MCMICHAEL, JOHN CALHOUN
APPLICANT: ZAGURSKY, ROBERT JOHN
APPLICANT: RUSSELL, DAVID PARRISH
APPLICANT: FLETCHER, LEAH DIANE
TITLE OF INVENTION: ALLOTOCOCUS OTITIDIS OPEN READING FRAMES (ORFs) ENCODING
FILE REFERENCE: AM100780 L2
CURRENT APPLICATION NUMBER: US/10/501,282
CURRENT FILING DATE: 2004-07-09
PRIOR APPLICATION NUMBER: 60/333,777
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 60/426,742
PRIOR FILING DATE: 2002-11-18

PRIOR APPLICATION NUMBER: PCT/US02/36123
 PRIOR FILING DATE: 2002-11-25
 NUMBER OF SEQ ID NOS: 6653
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 55
 LENGTH: 1212
 TYPE: DNA
 ORGANISM: *Allotococcus ocellidis*
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (463) ..(1209)
 US-10-501-282-55

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
 Best Local Similarity 46.7%; Pred. No. 7.7e-17;
 Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

184 TCATTGTGAGGAGATATGATGAGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGGG 243
 238 TCATTGTGAGGAGATATGATGAGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGGG 297
 244 GCAGACAGATATTTTCAATATGTTGAACGATCTTGAAGCTCGGATTATGACAGA 303
 298 TAGAGGAGATTTTACATATGTTAAACCAATTTGACGGTAGAGACTTGGTCGTC 357
 304 AACCTTGAACCCGGTAACTTATCAAAAGATTATA-----ACAAGA 348
 358 AATCTTGAATCTGCGCTACCTATGACAAATCGAATCTAGCTTAAGCCGCTTCAAGTTA 417
 349 GATTAAGATTCATCTGACAGCAATGAGATGAGTAAAGCTTGAAGATGATGAAT 408
 418 AGTGGAGTATTTATTTGATTTTTCATTTGACTGCTCAAGCCATGCAAGAGCTGGC 477
 409 TTCAAGGTTCTCAACAGCCCAACAACAACGATGATTAACGCGCTTCAAGGATGAAA 468
 478 ATTGACCTAGTTTCCATGCGCAACAACATACCGGGGACATGGGCAAGAGGATGGT 537
 469 GATACGCTTGAGAAATTTGGAGCAAAAACCTTGATATGTTGGAGCGGATACAGCTTA 528
 538 GACGGATGGAATTTTAAGAGATGTAAGATGATTAATTTGGAATGGGCCATGATCA 597
 529 AGTGAATGCAAAAAGAAATTTGCTACCAAGATGACCGGGTAAACGATGCAAGGCTT 588
 598 GCAGAAAGC---TGGCCAGCTTATCAATTTGTTGCAATTAATTTGACTTCAATTTT 654
 589 GGCTTAACGATGTGTCGGGAAAAGTTTCGCGCTTAAAGAAATACGCGCGGCTGCTG 648
 655 GCAGTTAGCATGTATTAACCTGGCAAGCAGCTAGTATGATGATGAACAGGATGATC 714
 649 CCGCAGATCTGAAATCTTCAATCCCTATGATTTCAAGAGCAAAAACATGCTGACAT 708
 715 AGCAACAATCCCAAGCTTTTCTTAACCTGGCCAAATGATATGACCAAGATCTGATTTG 774
 709 GTTGTGTGATGATCACTGGGGCAAGATATGACAAATATCCAAACGACCGCCAGCGC 768
 775 GTTATGCGCTATATTCATGCTGATGATGATTAATTCGTAACCTGATGAAAACACAG 834
 769 CAGCTTGCAAGAGCAATGCTGATGAGGGAGCTGACATCAATCGCGCCATCAACGAC 828
 835 GAGTGGCTGAAGCTTAATGATGCTGTGACAGATATTTGTTATTTGCTCCACACCCAC 894
 895 TCCTCTGCTGCTGTGAAAATATCCAAAGATGGGATATTTTATATGCTTGGGCAATTT 954
 889 GTCTTTGACCAAGCTTGACGAGAAACAGAGACATGCTGCTGCTGATCACTTAAG 948
 955 ATTTTGAACCAAGCATGCAAAAGTTCAACGATTCGCTTATTTTGAACATGATATCAAT 1014
 949 AAAAATGAAACAGCGCGCTTGAAGTACACCGATGATATCCATGAAGCAACACTGCA 1008
 1015 AGTCCAGACCAAGTTGCTTTTACCTTAAGGCCAATGAATAAGAGCTGGATTTCCCAA 1074

OY 1009 CC 1010
 DB 1075 CC 1076

RESULT 7
 US-10-501-282-57

Sequence 57, Application US/10501282
 Publication No. US20050203280A1

GENERAL INFORMATION:

APPLICANT: MCMICHAEL, JOHN CALHOUN

APPLICANT: ZAGURSKY, ROBERT JOHN

APPLICANT: RUSSELL, DAVID PARRISH

APPLICANT: FLETCHER, LEAH DIANE

TITLE OF INVENTION: ALLOTOCOCCUS OTTIDIS OPEN READING FRAMES (ORFS) ENCODING.

TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF

FILE REFERENCE: AM100780 U2

CURRENT APPLICATION NUMBER: US/10/501.282

CURRENT FILING DATE: 2004-07-09

PRIOR APPLICATION NUMBER: 60/333,777

PRIOR FILING DATE: 2001-11-29

PRIOR FILING DATE: 60/426,742

PRIOR FILING DATE: 2002-11-18

PRIOR APPLICATION NUMBER: PCT/US02/36123

PRIOR FILING DATE: 2002-11-25

NUMBER OF SEQ ID NOS: 6653

SOFTWARE: PatentIn version 3.2

SEQ ID NO 57

LENGTH: 1212

TYPE: DNA

ORGANISM: *Allotococcus ocellidis*

FEATURE:

NAME/KEY: CDS

LOCATION: (244) ..(1209)

US-10-501-282-57

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
 Best Local Similarity 46.7%; Pred. No. 7.7e-17;
 Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

184 TCATTGTGAGGAGATATGATGAGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGGG 243
 238 TCATTGTGAGGAGATATGATGAGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGGG 297
 244 GCAGACAGATATTTTCAATATGTTGAACGATCTTGAAGCTCGGATTATGACAGA 303
 298 TAGAGGAGATTTTACATATGTTAAACCAATTTGACGGTAGAGACTTGGTCGTC 357
 304 AACCTTGAACCCGGTAACTTATCAAAAGATTATA-----ACAAGA 348
 358 AATCTTGAATCTGCGCTACCTATGACAAATCGAATCTAGCTTAAGCCGCTTCAAGTTA 417
 349 GATTAAGATTCATCTGACAGCAATGAGATGAGTAAAGCTTGAAGATGATGAAT 408
 418 AGTGGAGTATTTATTTGATTTTTCATTTGACTGCTCAAGCCATGCAAGAGCTGGC 477
 409 TTCAAGGTTCTCAACAGCCCAACAACAACGATGATTAACGCGCTTCAAGGATGAAA 468
 478 ATTGACCTAGTTTCCATGCGCAACAACAACATTCGCGGACATGGGCAAGAGGATGGT 537
 469 GATACCTTGAGAAATTTGCGAAGCAAAAACCTGATATGTTGAGCGGAGATACAGCTTA 528
 538 GACGGATGGAATTTTAAAGATGTAAGATGATTAATTTGAAATGGGCCATGATGCA 597
 529 AGTGAATGCAAAAAGAAATTTGCTACCAAGAAAGTCAACGGGTTAACGATGCAAGCTT 588
 598 GCAGAAAGC---TGGCCAGCTTATCAATTTGTTGCAATTAATTTGATCACTTCAATTTT 654
 589 GGCTTAACGATGTGTCGGGAAAAGTTTCGCGCTTAAAGAAATACGCGCGGCTGCTG 648
 655 GCAGTTAGCATGTGATTAACCTGGCAAGCAGCTAGTATGATGATGAACAGGATGATC 714
 649 CCGCAGATCTGAAATCTTCAATCCCTATGATTTTGAAGAGCAAAAACATGCTGACAT 708

Db 715 ACGACCAACTCCCAAGCCTTTCTTAACCTGGCCAAATAGCTATGACCAAGAAATCTGATTGG 774
Qy 709 GTTGTGTGCACTACACTGCGGGCCCAAGATATGACAAATGATCCAAAGCCGCGAGGC 768
Db 775 GTTATCGCTTATATCCATGCTGGAATAGATATATTCCTCAACCTGATGCAAAACCCACAG 834
Qy 769 CAGCTTGCAAGAGCCATGCTGATGCGGAGCTGACATCATGTCGCGCATCATCCGAC 828
Db 835 GAGCTGGCTGAAAGCTTAAATGATGCTGTCAGATATTTGTTATTTGCTCCACACCCAC 894
Qy 829 GTCTTAGAACCGATTGAATATATACGAAACCGTATTTTCTACAGCCTCGCAACTTT 888
Db 895 TCCCTCCCTGCGTGTGAAATAATACCAAGATGGGATTTATTTTATGCTGGGCAATTTT 954
Qy 889 GTCTTTGACCAAGCTGGAACGAAACAGACAGTGCATGCTTCACTATCAGCTGAAG 948
Db 955 ATTTTGTACCAAGGATCAAAAGTTCAACCGATTCCTTATTTTATGACATGGAATATCAAT 1014
Qy 949 AAAATGGAACAGCGCGCTTTGAAGTGAACCGATGATTCATGAGACGACCTGCA 1008
Db 1015 AGTCCAGACCAAGGTTCTTTTACCTTAAGCCCAATGAAATGAAAGCTGATTTCCCAA 1074
Qy 1009 CC 1010
Db 1075 CC 1076

RESULT 8
US-10-501-282-59
; Sequence 59, Application US/10501282
; Publication No. US20050203280A1

GENERAL INFORMATION:
; APPLICANT: MONTICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 59
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)..(1209)
US-10-501-282-59

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
Best Local Similarity 46.7%; Pred. No. 7,7e-17;
Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

Qy 184 TCATTGTAGCGGATATATATGATGAGCCCTATGTTGAAAAAGTAAACGAGCAAAAAGGG 243
Db 238 TCATTGTAGCGGATATATATGATGAGCCCTATGTTGAAAAAGTAAACGAGCAAAAAGGG 297
Qy 244 GCAGCAGATTTTCAATATGATGTAACCGATCTTAGAGCTCGGATATATGACAGCA 303
Db 298 TACGAGGAATTTTACATATGTAACCAACATTTTGACGGTAGAGACTTGCTGCTGCC 357
Qy 304 AACTTGAAGAACCCGTAACCTATCAAAAAGATTAATACACAGCA 348

Db 358 AATCTGAATTCGCGCTACCTATGACAAATGAACTTACCTTAAGCCGCTCAAGTTTA 417
Qy 349 GATAAGAGATTCATCTGACAGCAATATGAAATCACTGAAAGTCTTGAAGATATGAAT 408
Db 418 AGTGAGGATTTTATTTGATTTCTTCACTGATGCTCAATGCAAGAAAGCTGGC 477
Qy 409 TTCAGGTTCTCAACAGCGCCAAACACACGCAATGATTAACGGCTTCAAGGGCATGAA 468
Db 478 ATTACCTAGTTTCATGCGCAACCAACATACCGGCACATGGGCAAGAGGGATGCTT 537
Qy 469 GATACGCTTGGAATTTGGAAGCAAAACCTTGATATCGTTGAGAGGGGATACAGTTA 528
Db 538 GACGCAATGAAATTTTAAAGATGTAAGATTAATATGAAATGGGCAATGATCA 597
Qy 529 AGTATGCAAAAAAGAAATTTCTGATACAGAAAGTCAACGGGGTAAAGATTCACAGCTT 588
Db 598 GCAGAGC---TGCCAGAGCTTATCAATTTGTCGAATTAATTTGACTTACTCATTTT 654
Qy 589 GCGTTTACCGATGTCGCGGAAAGTTTCCGCTTAAAGAAATACGCGGGGCTGCTG 648
Db 655 GCACTTAGCATGATTAATTAACCTGGCAAGAGCTAGTATGATGAACCAAGGTGATCTC 714
Qy 649 CCGCAGATCTGAAATCTTCAATCCCTATGATTTCAAGACGAAAAACATGCTGACAT 708
Db 715 ACGACCAACTCCCAAGCCTTTCTTAACCTGGCCAAATAGCTATGACCAAGATCTGATTTG 774
Qy 709 GTTGTGAGAGTCACTGCGGGCCAGAGATGATGACAAATGATCCAAACGACCGCAGCG 768
Db 775 GTTATCGCTTATATCCATGCTGGAATGATATATTTGCTCAACCTGATGCAACACACAG 834
Qy 769 CAGCTTGCAAGAGCATGCTGATGCGGAGCTGACATCATGTCGCGCATCATCCGAC 828
Db 835 GAGCTGGTGAAGACTTAATGATGCTGTCGACATATTTGTTATTTCTCCACACCCAC 894
Qy 829 GTCTTAGAACCGATTGAATATATACGAAACCTCATTTTCTACAGCTTGGCACTTT 888
Db 895 TCCCTCCCTGCTGTGAAATATCAAGATGGAATTTTATTTTATGCTTGGGCAATTTT 954
Qy 889 GTCTTAGAACAGGCTGGAAGCAACAGACAGTGCATGCTTCACTGATACCTGAAG 948
Db 955 ATTTTGAACCAAGCATGCAAAAGTTCAACCGATTCCTTATTTTACATGATATCAAT 1014
Qy 949 AAAATGGAACAGCGCGCTTTGAAGTGAACCGATGATTCATGAAAGCAACCTGCA 1008
Db 1015 AGTCCAGACCAAGGTTCTTTTACCTTAAGCCCAATGAAATGAAAGCTGATTTCCCAA 1074
Qy 1009 CC 1010
Db 1075 CC 1076

RESULT 9
US-10-501-282-6651/C
; Sequence 6651, Application US/10501282
; Publication No. US20050203280A1

GENERAL INFORMATION:
; APPLICANT: MONTICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2

SEQ ID NO 6651
 LENGTH: 1754382
 TYPE: DNA
 ORGANISM: *Alliostococcus oltididis*
 US-10-501-282-6651

Query Match 8.5%; Score 96.4; DB 9; Length 1754382;

Best Local Similarity 46.7%; Pred. No. 3.9e-15;
 Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

```

QY 184 TCATTGAGGCGAATATTATGATGGAGCGTATGTTGAAAAAGTAACGAGCAAAAAGG 243
DB 24408 TCATTGAGGCGAATATTATGATGGAGCGTATGTTGAAAAAGTAACGAGCAAAAAGG 24349
QY 244 GCAGACAGTATTTTCAATATGTTGAAACCGATCTTTAAGCCTCGAATTATGATGACAGA 303
DB 24348 TACGAGCAATTTTTCATATGTTAAACCAATTTTGACGGTAGAGACTGGTCGTTGCC 24289
QY 304 AACTTGAACCCCGTAACTATCAAAAGATTATA-----ACNAGCA 348
DB 24288 AATCTTGATCTGCCGTACCTATGACATCGAATCTAGTTAAGCCGCTTCAAGTTTA 24229
QY 349 GATAAGAGATCATCTGAGACGATAAGAAATAGAAATGAGTAAGTCTTGAAGGATATGAT 408
DB 24228 AGTGAAGTATTTATTTGATTTCTTCATGACTGATCAAGTCAAGGATGCAAGAGCTGGC 24169
QY 409 TTCAGGTTCTCAAGAGCGCCAAACCAACGCAATGATTAAGCGGCTTCAAGGCAATGAAA 468
DB 24168 ATTGACCTAGTTTCATGAGCCCAACCAATACCGGAGATGGGCAAGAGGATATGTT 24109
QY 469 GATACGCTTGAGAAATTTGCGAAGCAAAACCTTATATGTTGAGCGGAGATACAGCTTA 528
DB 24108 GACGCGATGAAATTTTAAAGATGTAGAAATGATATATTTATGATGGCCATGATGCA 24049
QY 529 AGTATGCGAATAAATAATTTCTACCAAGAAAGTCAAGGAGTATGCAAGCTT 588
DB 24048 GCAGAGAGC---TGCCAGAGCTATCAATTTGTTGCAATTAATTTGACTCTCCTCATTTT 23992
QY 589 GAGTTTACCGATGTGTCCGGGAAAGGTTTCGGGCTTAAAGAAATACGCGGCGGTGCTG 648
DB 23991 GCAGTTAGGATGTGATTAACCTGGCCAAAGCAGTATGATGATACCAAGGTACTC 23932
QY 649 CCCGAGATCTGAAATCTTCACTCTATGATTTTCAAGAACGAAAAAATAGTGTGACATT 708
DB 23931 ACGACCAACTCCCAAGCTTTCTTAACCTGSCCAATAGCTATGACCAAGAAATCTGATTTG 23872
QY 709 GTTGTGTGACATGACACTGGGGCCAAAGTATGACATGATCCAAAGCAAGCCGACGGC 768
DB 23871 GTTATGCGCTATATCTATGCTGGAATGATGATATTTGTCACCTGATGCAACCAAG 23812
QY 769 CAGCTTGCAAGAGCCATGTCATGTCGGGAGCTGATCATCTGCGGCATCTCCGAC 828
DB 23811 GAGCTGCTGAAAGCTTAATGATGCTGATGAGATTTGTTTGTCTCCCAACCCAC 23752
QY 829 GTCTTAGAACCGATTAATATTAACGAAACCGTCAATTTTCAAGCCTCGGCAACTTT 888
DB 23751 TCCCTCTGCTGTTTAAATAATCAAGATGGATATTTTATGAGCTTGCGCAATTTT 23692
QY 889 GTCTTTGACCAAGCTGAGAGAAACAGAGACAGTCACTGTTCAAGTACCTGAAAG 948
DB 23691 ATTTTGGACCAAGGCAATGCAACGATTCACGATTCGTTATTTTGAAGATGATATCAAT 23632
QY 949 AAAAAAGGAACAGCGCGCTTGAAGTACACGATGATTCATGAAAGGACACCTGCA 1008
DB 23631 AGTCCAGACCAAGGTTGTTTACCTTATGAGCAATGAAATATAGAGCTGATATCCCAA 23572
QY 1009 CC 1010
DB 23571 CC 23570

```

RESULT 10
 US-10-398-221-10/c

```

; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1163020
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
; US-10-398-221-10

```

Query Match 6.8%; Score 77.8; DB 7; Length 1163020;

Best Local Similarity 57.0%; Pred. No. 8.2e-10;
 Matches 142; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

QY 649 CCCGAGATCTCGAAATCTTCATCCCTATGATTCAGAAAGCAAAAACATGCTGACATT 708
DB 446758 CCAGCTATCTTGAAGAAATGTAAAGACGTTAAAAATATAAAGAAAGATACGCTC 446699
QY 709 GTTGTGTGACATGACACTGGGGCCAAAGATGACAAATGATCCAAAGACCGCCAGCGC 768
DB 446698 GTTATTTGTAATATGACACTGGGGCGCTCGAAATCTGAAACCTCCAAACCACTAAC 446639
QY 769 CAGCTTGCAAGAGCCATGCTGATGCGGAGCTGACATCATGCGGCATCATCCGAC 828
DB 446638 CAATTTGACATGATATTTTGAAGATGCTGATGATATTAATTTGAGAGCTATCCGAC 446579
QY 829 GTCTTAGAACCGATTAATATTAACGAAACCGTCAATTTTCTACAGCCTCGGCAACTTT 888
DB 446578 CCGCTGGAAGATGCAAAATATTAAGATTAATATTTGATATACAGATGCGCATTTT 446519
QY 889 GTCTTTGAC 897
DB 446518 GCTTTTGGC 446510

```

```

RESULT 11
US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
; US-10-398-221-2058

```


Query Match 6.8%; Score 77.8; DB 7; Length 3011208;
Best Local Similarity 57.0%; Pred. No. 1.4e-09;
Matches 142; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 649 CCCCAGATCTGAAATCTTCAATCCCTATGATTCAGAGGAGGAAAAATCATGCTGACAT 708
DB 537024 CCAGCCTATCTGTAAGAAATGTAAGAACGTTAAATAATTAAGAAAGATACGCTC 536965
OY 709 GTTGTTGGCAGTCACTGGGGCCCAAGATGACATGATCCAAAGACCGGAGCGC 768
DB 536964 GTTATGTTAATGACTGGGGCGTGAATATCGTGAACCTCCAACTACCAACT 536905
OY 769 CAGCTTGACAGAGCCATCTGATCGGAGCTGACATCATCTGCGCCATCATCCGAC 828
DB 536904 CAATTGACATGCTATTTAGATCTGCTGATATTAATTAAGAGCTCATCCAC 536845
OY 829 GTCTTGAACCGATTGAAGTATTAACGAAACCGTCAATTTCTACAGCCTCGCACTTT 888
DB 536844 CGCCTGAAAGTATCGAAATAATTAATTAATGATACAGTATGGCGCATTTT 536785
OY 889 GTCTTTGAC 897
DB 536784 GCTTTGGC 536776

RESULT 12

US-10-398-221-1526
; Sequence 1526, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1526
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Listeria monocytogenes-4B
US-10-398-221-1526

Query Match 6.5%; Score 74; DB 7; Length 885;
Best Local Similarity 54.4%; Pred. No. 2.2e-10;
Matches 149; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

OY 622 GCTAAAAAGATACGCGGGCGTCTGCCGAGATCCTGAATCTTATCCCTATGAT 681
DB 40 GCAAGACGACATACCAAGCGGATATCCATGCTGATTAAGAAATGTACGACAGTG 99
OY 682 TCAGAGGAGAAAAAATGCTGACATTTGTTGTGTCAGTCACTGCGGCGCAAGAT 741
DB 100 AAGACATACAAAAAAGAACTCTCTCGTTATCGATTAATACCACTGGGGCGTCAAT 159
OY 742 GACATGATCCAAAGACCGGAGCGGAGCTTGAAGAGCCATGCTGATCGGAGCT 801
DB 160 CGCGAAACCAACGAGATTTATTAACCCAAATTTGGTCATGCAATTTTATGATGCT 219
OY 802 GACATCATCTGCGGACATCATCCGACGCTTGAAGACGATGAAGTATTAACGAAAC 861
DB 220 GATATTAATATGAGGCTCTACACCTCATGCGCTAGAAAGGTTGAATAATTAAGAT 279
OY 862 GTCAATTTCTACAGCTCGGCACTTTGCTTTG 895
DB 280 TACATCGTTTATGATGAGGCGACTTTGCTTTG 313

Query Match 6.8%; Score 77.8; DB 7; Length 3011208;
Best Local Similarity 57.0%; Pred. No. 1.4e-09;
Matches 142; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 649 CCCCAGATCTGAAATCTTCAATCCCTATGATTCAGAGGAGGAAAAATCATGCTGACAT 708
DB 537024 CCAGCCTATCTGTAAGAAATGTAAGAACGTTAAATAATTAAGAAAGATACGCTC 536965
OY 709 GTTGTTGGCAGTCACTGGGGCCCAAGATGACATGATCCAAAGACCGGAGCGC 768
DB 536964 GTTATGTTAATGACTGGGGCGTGAATATCGTGAACCTCCAACTACCAACT 536905
OY 769 CAGCTTGACAGAGCCATCTGATCGGAGCTGACATCATCTGCGCCATCATCCGAC 828
DB 536904 CAATTGACATGCTATTTAGATCTGCTGATATTAATTAAGAGCTCATCCAC 536845
OY 829 GTCTTGAACCGATTGAAGTATTAACGAAACCGTCAATTTCTACAGCCTCGCACTTT 888
DB 536844 CGCCTGAAAGTATCGAAATAATTAATTAATGATACAGTATGGCGCATTTT 536785
OY 889 GTCTTTGAC 897
DB 536784 GCTTTGGC 536776

RESULT 13

US-10-398-221-3371
; Sequence 3371, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3371
; LENGTH: 2282
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3371

Query Match

6.5%; Score 74; DB 7; Length 2282;
Best Local Similarity 54.4%; Pred. No. 3.6e-10;
Matches 149; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

OY 622 GCTAAAAAGATACGCGGGCGTCTGCCGAGATCCTGAATCTTATCCCTATGAT 681
DB 40 GCAAGACGACATACCAAGCGGATATCCATGCTGATTAAGAAATGTACGACAGTG 99
OY 682 TCAGAGGAGAAAAAATGCTGACATTTGTTGTGTCAGTCACTGCGGCGCAAGAT 741
DB 100 AAGACATACAAAAAAGAACTCTCTCGTTATCGATTAATACGCACTGGGGCGTCAAT 159
OY 742 GACATGATCCAAAGACCGGAGCGGAGCTTGAAGAGCCATGCTGATCGGAGCT 801
DB 160 CGCGAAACCAACGAGATTTATTAACCCAAATTTGGTCATGATGATTTAGATGCTG 219
OY 802 GACATCATCTGCGGACATCATCCGACGCTTGAAGACGATGAAGTATTAACGAAAC 861
DB 220 GATTAATATGAGGCTCTACACCTCATGCGCTAGAAAGGTTGAATAATTAAGAT 279
OY 862 GTCAATTTCTACAGCTCGGCACTTTGCTTTG 895
DB 280 TACATCGTTTATGATGAGGCGACTTTGCTTTG 313

RESULT 14

US-10-470-048B-514
; Sequence 514, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; FILE REFERENCE: SONN:035US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; PRIOR FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 514
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-470-048B-514

Query Match 4.8%; Score 54.2; DB 8; Length 1071;
Best Local Similarity 47.4%; Pred. No. 0.00014;

Matches 229; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

QY 465 GAAAGATACGCTTGAGAAATTTGCGAAGCAAACTTGATATCGTTGGACGGGATACAG 524
 DB 432 GACAGCAATGATGAGAAAAATAGATTTCCAGACAGTTATATTTTAAACAGTAAATG 491
 QY 535 CTTAAGTATGCGAAAAAAATTTTCGACGAAAGTCAACGGGGTACAGATTGCAAC 584
 DB 492 TTCAAAATCCAAATTAATAGTAAACGTACACAGACATTAAGTAAAGTAAATAGCTAA 551
 QY 585 GCTTGCGCTTACGATGTGTCCGGGAAAGTTTCGGCGCTAAAGATACCGCGGCGT 644
 DB 552 TGTTCATTTACCGATATGAAATCTACATTAATCTTTAAAAAACAACAGC---TC 608
 QY 645 GCTGCCCGCAGATCTGGAATCTTCATCCCTATGATTTGAGAAGCGAAAAACATGCTGA 704
 DB 609 AATTAGTTAGATCCAGCTATATTTATCTTTAATAAAAAATTAAGAAAAATATATGA 668
 QY 705 CATTGTGTGTGCACTCACACTGGGGCCCAAGATGTACAAATGATCCAAACGACCGCA 764
 DB 669 TTACGTCTGATCAATGTAGATTTGGGGATACCTAATGACGAAATGTGACTACACATGA 728
 QY 765 GGGCCAGCTTGCAAGACCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCC 824
 DB 729 AAAAGATATGCAATGCTGACGATGCTGACGATGCTGACGATGCTGATGCTGATGCT 788
 QY 825 GCACGCTTTAGAACCGATTTGAATATTAACGAAACCGTATTTTCTACAGCTCGGCA 884
 DB 789 ---AGTTATTTCAAAAAGTTGAAAATATTAAGGAAAGCCATTTTATATAGTTAGTAA 845
 QY 885 CTTTGTCTTTGACCAAGCTGACGAGCAAGACAGACAGTGCACCTGCTTACATACCT 944
 DB 846 CACAACGCTGTAACTTTCTTATCAAAAATCAGAAAGAAATGATGTACACAGACTG 905
 QY 945 GAA 947
 DB 906 GAA 908

RESULT 15
 US-10-724-972A-1055
 ; Sequence 1055, Application US/10724972A
 ; Publication No. US20040147734A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Doucette-Stamm, Lynn
 ; APPLICANT: Bush, David
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: PATH03-16
 ; CURRENT APPLICATION NUMBER: US/10/724,972A
 ; CURRENT FILING DATE: 2003-12-01
 ; PRIOR APPLICATION NUMBER: 09/450,969
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: 09/134,001
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 7544
 ; SEQ ID NO 1055
 ; LENGTH: 1101
 ; TYPE: DNA
 ; ORGANISM: S. epidermidis
 US-10-724-972A-1055

Query Match 4.8%; Score 54.2; DB 7; Length 1101;
 Best Local Similarity 47.4%; Pred. No. 0.00014;
 Matches 229; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

QY 465 GAAAGATACGCTTGAGAAATTTGCGAAGCAAACTTGATATCGTTGGACGGGATACAG 524
 DB 459 GACAGCAATGATGAGAAAAATAGATTTCCAGACAGTTATATTTTAAACAGGTAATG 518

QY 525 CTTAAGTATGCGAAAAAAGAAATTTCTGACGAAAGTCAACGGGGTACGATTGCAAC 584
 DB 519 TTCAAAATCCAAATTAATAGTAAACGTACACAGACATTAAGGTAAATAAATAGCTAA 578
 QY 585 GCTTGCGCTTACGATGTGTCCGGGAAAGTTTCGGCGCTAAAGATACCGCGGCGT 644
 DB 579 TGTTCATTTACCGATATGAAATCTACATTAATCTTTAAAAAACAACAGC---TC 635
 QY 645 GCTGCCCGCAGATCTGGAATCTTCATCCCTATGATTTGAGAAGCGAAAAACATGCTGA 704
 DB 636 AATTAGTTAGATCCAGCTATATTTATCTTTAATAAAAAATTAAGAAAAATATATGA 695
 QY 705 CATTGTGTGTGCACTCACACTGGGGCCCAAGATGTACAAATGATCCAAACGACCGCA 764
 DB 696 TTACGTCTGATCAATGTAGATTTGGGGATACCTAATGACGAAATGTGACTACACATGA 755
 QY 765 GGGCCAGCTTGCAAGACCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCC 824
 DB 756 AAAAGATATGCAATGCTGACGATGCTGACGATGCTGACGATGCTGATGCTGATGCT 815
 QY 825 GCACGCTTTAGAACCGATTTGAATATTAACGAAACCGTATTTTCTACAGCTCGGCA 884
 DB 816 ---AGTTATTTCAAAAAGTTGAAAATATTAAGGAAAGCCATTTTATATAGTTAGTAA 872
 QY 885 CTTTGTCTTTGACCAAGCTGACGAGCAAGACAGACAGTGCACCTGCTTACATACCT 944
 DB 873 CACAACGCTGTAACTTTCTTATCAAAAATCAGAAAGAAATGATGTACACAGACTG 932
 QY 945 GAA 947
 DB 933 GAA 935

Search completed: February 27, 2006, 07:58:43
 Job time : 1259.33 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 686.994 Seconds
(without alignments)
3538.457 Million cell updates/sec

Title: US-10-789-164-3

Perfect score: 1140

Sequence: 1 atgaaaaaagaactgagctt.....gtgacaactaataatctaa.1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA.New.*
1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.1.*
7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.1.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.1.*
11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.3.*
12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.4.*
13: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.5.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1138.4	99.9	1140	US-10-530-083-3	Sequence 3, Appl
2	47.6	4.2	3523	US-10-793-626-4412	Sequence 4412, Ap
3	42.4	3.7	173120	US-11-114-798-55	Sequence 55, Appl
4	42.4	3.7	182303	US-11-121-086-45	Sequence 53, Appl
5	42.4	3.7	196716	US-11-114-798-53	Sequence 474, App
6	41.8	3.7	3256	US-11-072-512-474	Sequence 700467, App
7	39	3.4	620	US-09-925-065A-700467	Sequence 700468, App
8	39	3.4	620	US-09-925-065A-55305	Sequence 55305, A
9	38	3.3	549	US-09-925-065A-73470	Sequence 5053, Ap
10	37.6	3.3	771	US-10-467-657-5053	Sequence 5055, Ap
11	37.6	3.3	1065	US-10-467-657-5055	Sequence 144642, Ap
12	36.6	3.2	543	US-09-925-065A-144642	Sequence 901662, A
13	36.6	3.2	612	US-09-925-065A-901662	Sequence 73465, A
14	36.2	3.2	745	US-09-925-065A-73465	Sequence 73470, A
15	36.2	3.2	745	US-09-925-065A-73470	Sequence 73471, A
16	36.2	3.2	1953	US-11-128-660-2	Sequence 2, Appl
17	36.2	3.2	730	US-09-925-065A-84661	Sequence 84661, A
18	35.8	3.1	50959	US-11-117-187-210	Sequence 210, App
19	35.8	3.1	50959	US-11-117-187-210	Sequence 206, App
20	35.8	3.1	72600	US-11-117-187-206	Sequence 206, App

21	35.4	3.1	575	6	US-09-925-065A-179132	Sequence 179132, A
22	35.2	3.1	578	6	US-09-925-065A-370910	Sequence 370910, A
23	35.2	3.1	623	6	US-09-925-065A-329381	Sequence 329381, A
24	35.2	3.1	623	6	US-09-925-065A-329382	Sequence 329382, A
25	35	3.1	574	6	US-09-925-065A-591106	Sequence 591106, A
26	35	3.1	574	6	US-09-925-065A-591107	Sequence 591107, A
27	34.8	3.1	616	6	US-09-925-065A-577593	Sequence 577593, A
28	34.6	3.0	636	6	US-09-925-065A-727350	Sequence 727350, A
29	34.6	3.0	3205	6	US-09-925-065A-698972	Sequence 698972, A
30	34.6	3.0	3205	6	US-09-925-065A-615393	Sequence 615393, A
31	34.4	3.0	577	6	US-09-925-065A-440264	Sequence 440264, A
32	34.2	3.0	581	6	US-09-925-065A-422278	Sequence 422278, A
33	34.2	3.0	621	6	US-10-750-185-34362	Sequence 34362, A
34	34.2	3.0	1976	8	US-10-750-623-34362	Sequence 34362, A
35	34.2	3.0	98345	12	US-11-112-908-36	Sequence 36, Appl
36	34.2	3.0	498	8	US-10-793-626-2527	Sequence 2527, Ap
37	34	3.0	578	6	US-09-925-065A-370909	Sequence 370909, A
38	34	3.0	618	6	US-09-925-065A-241898	Sequence 241898, A
39	34	3.0	678	6	US-09-925-065A-853865	Sequence 853865, A
40	34	3.0	3504	8	US-10-793-626-3859	Sequence 3859, Ap
41	34	3.0	5010	12	US-11-110-480-46	Sequence 46, Appl
42	34	3.0	5012	12	US-11-110-480-58	Sequence 58, Appl
43	34	3.0	5025	12	US-11-110-480-26	Sequence 26, Appl
44	34	3.0	5037	12	US-11-110-480-36	Sequence 36, Appl
45	34	3.0	5037	12	US-11-110-480-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-10-530-083-3
Sequence 3, Application US/10530083
General Information:
US20050249752A1
Applicant: Sung, Moon-Hee
Applicant: Lee, Jong-Seo
Applicant: Uung, Chang-Min
Applicant: Hong, Seong-Pyo
Applicant: Kim, Chul-Joong
Applicant: Park, Sue-nie
Applicant: Pyo, Hyun-mi
Title of Invention: VECTOR FOR ANTI-HPV VACCINE AND TRANSFORMED MICROORGANISM BY THE
File Reference: 4240-119
Current Application Number: US/10/530,083
Current Filing Date: 2005-04-01
Prior Application Number: KR 10-2002-0063378
Prior Filing Date: 2002-10-17
Number of Seq ID NOS: 11
Software: PatentIn version 3.2
Seq ID NO 3
Length: 1140
Type: DNA
Organism: Bacillus subtilis
US-10-530-083-3
Query Match 99.9%; Score 1138.4; DB 8; Length 1140;
Best Local Similarity 99.9%; Pred. No. 6.2e-300;
Matches 1139; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAAAAGAACTGCTTTCATGAAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
DB 1 ATGAAAAAAGAACTGCTTTCATGAAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
QY 61 AAAACCAATAGACGATTTTATGCAATTCGATGCTTTTGTCTTATGTCGCTTTC 120
DB 61 AAAACCAATAGACGATTTTATGCAATTCGATGCTTTTGTCTTATGTCGCTTTC 120
QY 121 ATGTGGCGGGAAGAAAGCGGAAAGCGGAAAGTCAAAAGCTATTCGACGAGTACTCTCA 180
DB 121 ATGTGGCGGGAAGAAAGCGGAAAGCGGAAAGTCAAAAGCTATTCGACGAGTACTCTCA 180

```

OY 181 GCGTCATTGTAGGCGATATTATGATGAGGAGCTATGTTGAAAAAGTAACGAGCAAAA 240
    |||
Db 181 GCGTCATTGTAGGCGATATTATGATGAGGAGCTATGTTGAAAAAGTAACGAGCAAAA 240
OY 241 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGAGCCTCGGATTTATGTACA 300
    |||
Db 241 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGAGCCTCGGATTTATGTACA 300
OY 301 GGAACCTTTGAAAAACCCGGTAACTTATCAAAAAGAAATTTAAACAAGCAGATTAAGAGTT 360
    |||
Db 301 GGAACCTTTGAAAAACCCGGTAACTTATCAAAAAGAAATTTAAACAAGCAGATTAAGAGTT 360
OY 361 CATCTGCGAGAGATTAAGAAATCAAGTAAAGCTTGAAGAGATATGAAATTTACCGTTCTC 420
    |||
Db 361 CATCTGCGAGAGATTAAGAAATCAAGTAAAGCTTGAAGAGATATGAAATTTACCGTTCTC 420
OY 421 AACAGCGCCAAACAACGCAATGATTAACGCGTTCAAGGCGATGAAAGATACGTTGGA 480
    |||
Db 421 AACAGCGCCAAACAACGCAATGATTAACGCGTTCAAGGCGATGAAAGATACGTTGGA 480
OY 481 GAATTTGGGAAAGCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAGTATGCCAAA 540
    |||
Db 481 GAATTTGGGAAAGCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAGTATGCCAAA 540
OY 541 AAGAAAAATTTGTTACAGAAAGTCAACGAGGAGTAAAGATTTGCAACGCTTGGCTTACCGAT 600
    |||
Db 541 AAGAAAAATTTGTTACAGAAAGTCAACGAGGAGTAAAGATTTGCAACGCTTGGCTTACCGAT 600
OY 601 GTGTCCGGGAAAAGTTTGGCGGCTAAAAAGAAATACGCCGGGGGTGTCTGCCGCAATCCT 660
    |||
Db 601 GTGTCCGGGAAAAGTTTGGCGGCTAAAAAGAAATACGCCGGGGGTGTCTGCCGCAATCCT 660
OY 661 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGATGATTTGTTGTGAG 720
    |||
Db 661 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGATGATTTGTTGTGAG 720
OY 721 TCACACTGGGGCCAGAGTATGACATGATCCAAACGACCCGACGCGCACTTGCAAGA 780
    |||
Db 721 TCACACTGGGGCCAGAGTATGACATGATCCAAACGACCCGACGCGCACTTGCAAGA 780
OY 781 GCCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCCGACGCTTTAGAACG 840
    |||
Db 781 GCCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCCGACGCTTTAGAACG 840
OY 841 ATGGAATATATTAACGAAACCGTCAATTTTCTACAGCTCGGCACTTTGTCTTTGACCA 900
    |||
Db 841 ATGGAATATATTAACGAAACCGTCAATTTTCTACAGCTCGGCACTTTGTCTTTGACCA 900
OY 901 GGGTGGAGAGAAACAAGACAGTGCACGTGTTCAAGTATCACTGAAGAAAAATGAAACA 960
    |||
Db 901 GGGTGGAGAGAAACAAGACAGTGCACGTGTTCAAGTATCACTGAAGAAAAATGAAACA 960
OY 961 GGCCGCTTTGAAGTACACCGATCGATATCCATGAAGGACACCTGCACTGTGAAAAAA 1020
    |||
Db 961 GGCCGCTTTGAAGTACACCGATCGATATCCATGAAGGACACCTGCACTGTGAAAAAA 1020
OY 1021 GACAGCCTTTGAAGTACACCGATCGATATCCATGAAGGACACCTGCACTGTGAAAAAA 1080
    |||
Db 1021 GACAGCCTTTGAAGTACACCGATCGATATCCATGAAGGACACCTGCACTGTGAAAAAA 1080
OY 1081 AAAGTAGAAGACGGAACCTGAGTGTGATTTGATGATGATGATGATGATGATGATGATGAT 1140
    |||
Db 1081 AAAGTAGAAGACGGAACCTGAGTGTGATTTGATGATGATGATGATGATGATGATGATGAT 1140

```

```

RESULT 2
US-10-793-626-4412
; Sequence 4412, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

```

```

; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4412
; LENGTH: 3523
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-4412

```

```

Query Match 4.2%; Score 47.6; DB 8; Length 3523;
Best Local Similarity 54.6%; Pred. No. 0.016;
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```

```

OY 686 AAGCGAAAAACATGCTGACATTTGTTGTCAGTCACTGGGGCCAGATATGACA 745
    |||
Db 2635 AAGCGAAAAACATGCTGACATTTGTTGTCAGTCACTGGGGCCAGATATGACA 2634
OY 746 ATGATCCAAACGACCGCCAGCCAGCTTGCAAGAGCCATGTCTGATCGGAGCTGACA 805
    |||
Db 2695 ATGATCCAAACGACCGCCAGCCAGCTTGCAAGAGCCATGTCTGATCGGAGCTGACA 804
OY 806 TCATGCTGGCCATCATTCGCGACGCTTTAGAACCGATTTGAAGTATTAACGGA 859
    |||
Db 2755 TAGTTTAAAGTACATCTCATGATGATTAACCAAGTAAGGTTGATGATGATGATGATGATGAT 2808

```

```

RESULT 3
US-11-114-798-55/C
; Sequence 55, Application US/11114798
; Publication No. US20060035246A1
; GENERAL INFORMATION:
; APPLICANT: WU, RINA
; APPLICANT: MARQUEZ, ABBEY
; TITLE OF INVENTION: CHROMOGENIC IN SITU HYBRIDIZATION METHODS, KITS, AND
; FILE REFERENCE: 0618.011.0004
; CURRENT APPLICATION NUMBER: US/11/114,798
; CURRENT FILING DATE: 2005-04-26
; PRIOR APPLICATION NUMBER: 10/173,525
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/952,851
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 55
; LENGTH: 173120
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-11-114-798-55

```

```

Query Match 3.7%; Score 42.4; DB 9; Length 173120;
Best Local Similarity 50.5%; Pred. No. 2.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
OY 368 AGACGATAGGAATCACTGAAAGCTTTGAAAGATATGATTTTACGCTTCTCAACGCG 427
    |||
Db 156108 AAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 156049
OY 428 CCAACAAACGACGATGATTAACGCGCTTCAAGGCAATGAAGATACGCTTGAAGATTG 487
    |||
Db 156048 AAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 155989
OY 488 CGAAGCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAGTATGCAAAAAAGAAA 547
    |||

```

DB 155988 AAAAGAAAAAGAAAAAGAAAGCAAGATAGACCCAAAGGCGATGAAATTGAAAG 155929

QY 548 TTTCGTACCAAGAACTCAACGGGG 571
DB 155928 CTTCCCTCCAGAAAGCCAGCCGAG 155905

RESULT 4
US-11-121-086-45/c

Sequence 45, Application US/11121086
Publication No. US2005026459A1
GENERAL INFORMATION:
APPLICANT: POULSEN, TIM S.
APPLICANT: NIELSEN, KIRSTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09138, 6000-00000
CURRENT APPLICATION NUMBER: US/11/121, 086
CURRENT FILING DATE: 2005-05-04
PRIOR APPLICATION NUMBER: 60/567,570
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 45
LENGTH: 182303
TYPE: DNA
ORGANISM: Homo sapiens
US-11-121-086-45

Query Match
Best Local Similarity 50.5%; Pred. No. 2.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Score 42.4; DB 12; Length 182303;

QY 368 AGACGATTAAGATTCAGTGAAGCTTGAAGATATGATTTCAAGCTTCAACAGCG 427
DB 178022 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 177963
QY 428 CCAACACACCGCATGATTAAGCGGCTTCAAGGCGATGAAGATACGCTTGAGAAATTGG 487
DB 177962 AAAAGAAAGAAAGAAAGAAAGAAAGCAATAGCCAGAACTTGTAGAAAGAAAGAAAGAA 177903
QY 488 CGAAGCAAAACCTGTATATCTTGGAGCGGAGATACGCTTGAAGAAATTGG 547
DB 177902 AAAAGAAAGAAAGAAAGAAAGAAAGCAAGATTAAGCCAAAGGCGAGTGAATTAAGAAAG 177843
QY 548 TTTCGTACCAAGAACTCAACGGGG 571
DB 177842 CTTCCCTCCAGAAAGCCAGCCGAG 177819

RESULT 5
US-11-114-798-53/c

Sequence 53, Application US/11114798
Publication No. US20060035246A1
GENERAL INFORMATION:
APPLICANT: WU, RINA
APPLICANT: MAROEZ, ABBEY
TITLE OF INVENTION: CHROMOSOMIC IN SITU HYBRIDIZATION METHODS, KITS, AND
FILE REFERENCE: 0618, 011,0004
CURRENT APPLICATION NUMBER: US/11/114, 798
CURRENT FILING DATE: 2005-04-26
PRIOR APPLICATION NUMBER: 10/173,525
PRIOR FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: 09/952,851
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn Ver. 3.3
SEQ ID NO 53
LENGTH: 196716
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: nucleotide sequence
US-11-114-798-53

Query Match
Best Local Similarity 50.5%; Pred. No. 2.5;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 368 AGACGATTAAGATTCAGTGAAGCTTGAAGATATGATTTCAAGCTTCAACAGCG 427
DB 13472 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 13413
QY 428 CCAACACACCGCATGATTAAGCGGCTTCAAGGCGATGAAGATACGCTTGAGAAATTGG 487
DB 13412 AAAAGAAAGAAAGAAAGAAAGAAAGCAATAGCCAGAACTTGTAGAAAGAAAGAAAGAA 13353
QY 488 CGAAGCAAAACCTGTATATCTTGGAGCGGAGTACGCTTGAAGATACGAGAAAGAAAGAA 547
DB 13352 AAAAGAAAGAAAGAAAGAAAGAAAGCAAGATTAAGCCAAAGGCGAGTGAATTAAGAAAG 13293
QY 548 TTTCGTACCAAGAACTCAACGGGG 571
DB 13292 CTTCCCTCCAGAAAGCCAGCCGAG 13269

RESULT 6
US-11-072-512-474

Sequence 474, Application US/11072512
Publication No. US20060029945A1
GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHITO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOKYUKI
APPLICANT: NAGAHARA, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: Novel full length cDNA
FILE REFERENCE: 084335-0191
CURRENT APPLICATION NUMBER: US/11/072,512
CURRENT FILING DATE: 2005-03-07
PRIOR APPLICATION NUMBER: US 60/350,978
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: JP 2001-379298
PRIOR FILING DATE: 2001-11-05
NUMBER OF SEQ ID NOS: 4096
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 474
LENGTH: 3256
TYPE: DNA
ORGANISM: Homo sapiens

US-11-072-512-474

Query Match
Best Local Similarity 50.2%; Pred. No. 0.59;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 243 GCGACAGATATTTTCAATATATGTTGACCAAGCTTTAAGCGCTCGAGTTATGATGACAG 302
DB 1684 GCGAAATAGTGTCTCATGATGATGAGCAGAGCTTGAAGCTTGTATTAAGAAAGCAG 1743
QY 303 AAACCTTGAAGAAAGCCGATTAACATTAAGAAAGATTAAGCAAGAGATTAAGATTA 362

Db 1744 TCTCTGTGTCCTTGGCGAGATCTCTGAAAACAGCCTTAACCTCCCTGATGTCACAGGTGGC 180

QY 363 TCTTCAGACGATAGAAGATCAGTGAAGTCTTGAAGATATGAATTTCAAGTTCGAA 422

Db 1804 TGGACTGGTGAAGATTACTCTGAAGATTTCTTGACGCAATGAATGATATCAGACCAG 1863

QY 423 CAGGCGCCAAACCAACGCAATGAT 447

Db 1864 TGCCACGAGGGAATAGCAATTGAT 1888

```

RESULT 7
US-09-925-065A-700467/c
; Sequence 700467, Application US/09925065A
; Publication No. US2004018104A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700467
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-700467

```

```

Query Match Similarity 3.4%; Score 39; DB 6; Length 620;
Best Local Similarity 49.3%; Pred. No. 1.6;
Matches 102; Conservative 0; Mismatches 105; Indels 0; Gaps 0

OY      225  ACTACGAGCAAAAAGGGGACAGACAGTATTTTTCATATGTTGAACCCATCTTTAGAC 284
Db      370  ATTTATGTCGCCACATTTAGTGGGGCATGTAAATGATGGGGATACCTGACCCCTTGTC 311
OY      285  CTCGATTATGTAGCAGGAAACCTTTGMAAACCCGGTAACTCTATCAAAAGATTATAACA 344
Db      310  AGGGATGTTTCCACAACATCCTTGATTAACCTGTCCCGATCTGTTGGTCTTAATC 251
OY      345  AGCAATTAAGAGATTTCATCTGCACAGCAATTAAGAAATCAGTAAAGCTTTGAAGATAT 404
Db      250  CATTAATGATGGGGTTGACATGGGAGGTATCATGATAGATGAGATTGCAACAAGATGT 191
OY      405  GAATTTCACGGTTCTCAACAGCGCCAA 431
Db      190  GTACAATCGGGGACATCATGCGCAA 164

RESULT 8
US-09-925-065A-700468/C
; Sequence 700468, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08

```

```

; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700468
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-700468

Query Match      3.4%; Score 39; DB 6; Length 620;
Best Local Similarity 49.3%; Pred. No. 1.6;
Matches 102; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Dn      225 AGTACGAGCAAAAAGGGGCGACAGATATTTTCAATATGTGTAACGATCTTTAGAGC 284
      |||||
Dn      370 ATTATGTCCCAATTAGTGGGCAATGTAPATGATGGGATACATGACCTTTGTCTC 311
      |||||
Dn      285 CTCGATTTATGTACGACGAAACITTTGAAAAACCCGGTAACCTATCAAAAAGATTATTAACA 344
      |||||
Dn      310 AGGGGATGTTTCCACAAACATCTTTGGATTAACCTGTCTCCGATCTGTTGGTCTTAATC 251
      |||||
Dn      345 AGCAGATPAAGAGATTTCATCTGCACAGACGAATTAAGAAATCACTGAAGCTTTGAAGATAT 404
      |||||
Dn      250 CATTAATGATGGGGTGTGAGCATGGAGAGTATCACTAGATAGATTAAGCAAAACAGAGATG 191
      |||||
Dn      405 GAATTTACCGSTTCTCAACAGCGCCA 431
      |||||
Dn      190 GTACACCTCGGGGCACATCAGGCCAA 164
      |||||

```

```

RESULT 9
US-09-925-065A-55305/C
/ Sequence 55305, Application US/09925065A
/ Publication No. US2004018104B1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ FILE REFERENCE: 108827.135
/ CURRENT APPLICATION NUMBER: US/09/925, 065A
/ CURRENT FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243, 096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250, 092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261, 766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289, 846
/ PRIOR FILING DATE: 2001-05-09
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 55305
/ LENGTH: 549
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-925-065A-55305

Query Match          3.3%; Score 38; DB 6; Length 549;
Best Local Similarity 59.1%; Pred. No. 2.9;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

1 ATGAAAAAGACACTGAGCTTTCATGAAAAAGCTGCTAAAGCTGACAAAAACACAAAAAAG 60

```


Db 187 ATGAATCCGACTGAGTGTGATGAGCCTTACTTACGTGAGTTACAGTATATATTA 128
Qy 61 AAACCAATAAGACGATTTATTTGCGATTCGATCGTTTGTCTTAT 110
Db 127 AAAACAGAGAAAGATTTTCTTTCTTTCTTTCTTTTCTTTT 78

RESULT 10

US-10-467-657-5053
; Sequence 5053, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACT Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMan99, version 1.04
; SEQ ID NO 5053
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5053

Query Match 3.3%; Score 37.6; DB 8; Length 771;
Best Local Similarity 56.5%; Pred. No. 4.3;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 768 CCAGCTTGCAAGACCATCTGATCGGAGCTGACATCGTGGCCATCATCCGA 827
Db 426 CGAGTTTGGCGGCTCGCGTGCATTCGCGCGGATGTCGATTCGCGGCGGCGCA 485
Qy 828 CGCTTGAACCGATTTGATATATACGGAACCGCATTTTCTACAGCTGGCACTT 887
Db 486 CGTTACTCGCGCGTTCGACCTTACCGACGCGCTTCACTCTTACAGCGGCGCACTT 545
Qy 888 TGTC 891
Db 546 TGCC 549

RESULT 11

US-10-467-657-5055/C
; Sequence 5055, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACT Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMan99, version 1.04
; SEQ ID NO 5055
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5055

Query Match 3.3%; Score 37.6; DB 8; Length 1065;
Best Local Similarity 56.5%; Pred. No. 5;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 768 CCAGCTTGCAAGACCATCTGATCGGAGCTGACATCATCTCGGCCATCATCCGA 827
Db 448 CGAGTTTGGCGGCTCGCGTGCATTCGCGCGGATGTCGATTCGCGGCGGCGCA 389
Qy 828 CGCTTGAACCGATTTGATATATACGGAACCGCATTTTCTACAGCTGGCACTT 887
Db 388 CGTTACTCGCGCGTTCGACCTTACCGACCGCTTCACTCTTACAGCGGCGCACTT 329
Qy 888 TGTC 891
Db 328 TGCC 325

RESULT 12

US-09-925-065A-144642/C
; Sequence 144642, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144642
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-144642

Query Match 3.2%; Score 36.6; DB 6; Length 543;
Best Local Similarity 52.3%; Pred. No. 6.9;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 412 ACGTTTCTCAACAGCGCCAAACACGATGATTAAGCGGCTTACGAGCATGAAAGT 471
Db 301 AAGATTTCAGAAAATTAATAAAGATTAAGATCAATATACAAAGACTGTAAAT 242
Qy 472 ACGCTTGAGATTTGGCAAGCAAAACCTGATTCGTTGGAGGGGATACGTTAAGT 531
Db 241 AGTAATATAGATGACATATATCAAAATCTTATCAATTTAGTAAGAAAGAAATGAAT 182
Qy 532 GATGCGAAAAGAAAATTTCTACCGAAGTCAA 566
Db 181 GATGTGAAAATTAAGTTGTGATATGAAAAATTTAA 147

RESULT 13

US-09-925-065A-901662/C
; Sequence 901662, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A

```

; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 901662
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-901662
```

```
Query Match 3.2% Score 36.2; DB 6; Length 612;
Best Local Similarity 52.3%; Pred. No. 7.3;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
```

```
QY 412 ACGGTTCTCAACAGCCGCAACGCAATGATTCGGCGTTACGGCGATGAAAGAT 471
DB 372 AAGATTCTGAAATAATAATAATAAAGATTAGATGATATATCAAGAGCTGTTAAT 313
QY 472 ACGCTTGAGAAATTTGCCAAGCAAACTTGATATCGTTGAGCGGGATACGCTTAAGT 531
DB 312 AGTAATATGAGATGACATATATCAATCTTATCAATTTAGTAAAGAAAGAAATGAAT 253
QY 532 GATGCGAATAAGAAATTTGATACGAGAAAGTCA 566
DB 252 GATGGAATAATTAAGTTGTGATGAAAAAATTA 218
```

```
RESULT 14
US-09-925-065A-73469/C
; Sequence 73469, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73469
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-73469
```

```
Query Match 3.2% Score 36.2; DB 6; Length 745;
Best Local Similarity 56.2%; Pred. No. 10;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```
QY 250 AGATATTTTCATATGTTGAACCGATCTTTAGAGCCTCGGATTAATGACGAAACTTT 309
DB 437 AGTTTTCAGATTTCTTGAAATAATAATTTCTTCACTGCTTATGCTGTTAGAGCCTTT 378
```

```
QY 310 GAAACCCGGTAACCTATCAAAAAGATTATTAACACAGATTAAGAGATTCACTGCGAG 369
DB 377 CCAAAACCTGTAATTTTCAAAAATTAATTTTCACTGCTCATGAGGCGCTGAGATTCACTG 318
QY 370 A 370
DB 317 A 317
```

```
RESULT 15
US-09-925-065A-73470/C
; Sequence 73470, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73470
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-73470
```

```
Query Match 3.2% Score 36.2; DB 6; Length 745;
Best Local Similarity 56.2%; Pred. No. 10;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```
QY 250 AGATATTTTCATATGTTGAACCGATCTTTAGAGCCTCGGATTAATGACGAAACTTT 309
DB 437 AGTTTTCAGATTTCTTGAAATAATAATTTCTTCACTGCTTATGCTGTTAGAGCCTTT 378
QY 310 GAAACCCGGTAACCTATCAAAAAGATTATTAACACAGATTAAGAGATTCACTGCGAG 369
DB 377 CCAAAACCTGTAATTTTCAAAAATTAATTTTCACTGCTCATGAGGCGCTGAGATTCACTG 318
QY 370 A 370
DB 317 A 317
```

Search completed: February 27, 2006, 11:38:28
Job time : 686.994 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 268,258 Seconds
(without alignments)
7553,984 Million cell updates/sec

Title: US-10-789-164-3

Perfect score: 1140
Sequence: 1 atgaaaaaagacgcagctt.....gtgacaactaaatcctaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/1/COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5/COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A/COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B/COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H/COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PC/COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP/COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE/COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	55.8	4.9	7218	2	US-08-232-463-14
C 2	54.2	4.8	1101	3	US-09-134-001C-1575
C 3	47.8	4.2	939	3	US-09-902-540-6958
C 4	47.8	4.2	4604	3	US-09-902-540-602
C 5	47.6	4.2	1170	3	US-09-134-001C-1479
C 6	47.6	4.2	3523	3	US-09-710-279-4412
C 7	47.2	4.1	536165	3	US-09-214-808-1
C 8	45	3.9	1116	3	US-09-902-540-3551
C 9	45	3.9	17228	3	US-09-902-540-1170
C 10	42.4	3.7	14554	3	US-09-949-016-15908
C 11	41.8	3.7	3256	3	US-10-104-047-474
C 12	39.4	3.5	1272	3	US-09-902-540-9266
C 13	39.4	3.5	9146	3	US-09-902-540-990
C 14	39	3.4	1785	3	US-09-328-352-1364
C 15	37.4	3.3	378	3	US-09-583-110-789
C 16	37.4	3.3	483	3	US-09-583-110-788
C 17	37.4	3.3	606	3	US-09-107-433-634
C 18	37.2	3.3	1107	3	US-09-902-540-8213
C 19	37.2	3.3	1992	3	US-09-107-532A-3207
C 20	37.2	3.3	12183	3	US-09-902-540-1065
C 21	37.2	3.3	50563	3	US-09-949-016-15821
C 22	36.8	3.2	9048	3	US-08-961-527-159
C 23	36.6	3.2	1141	3	US-09-806-708B-22
C 24	36	3.2	399	3	US-09-621-976-8976

25	35.8	3.1	860	3	US-08-858-207A-172	Sequence 172, App
26	35.6	3.1	1141	3	US-09-806-708B-22	Sequence 22, Appl
27	35.4	3.1	1664976	3	US-08-916-421B-1	Sequence 1, Appl1
28	35.4	3.1	1664976	3	US-09-692-570-1	Sequence 1, Appl1
C 29	35.2	3.1	118382	3	US-09-949-016-15996	Sequence 15996, A
C 30	35.2	3.1	118382	3	US-09-949-016-15997	Sequence 15997, A
C 31	35	3.1	1245	3	US-09-248-796A-1300	Sequence 1300, Ap
C 32	34.6	3.0	4121	3	US-09-638-524A-1	Sequence 1, Appl1
C 33	34.6	3.0	106746	3	US-09-326-402C-12	Sequence 12, Appl1
C 34	34.6	3.0	106746	3	US-08-832-867-1	Sequence 1, Appl1
C 35	34.4	3.0	4137	3	US-09-949-016-13300	Sequence 13300, A
C 36	34.4	3.0	75176	3	US-09-949-016-18033	Sequence 18033, A
C 37	34.2	3.0	474	3	US-09-621-976-18033	Sequence 16661, A
C 38	34.2	3.0	81819	3	US-09-949-016-16661	Sequence 16662, A
C 39	34.2	3.0	81819	3	US-09-949-016-16662	Sequence 16662, A
C 40	34	3.0	498	3	US-09-710-279-2527	Sequence 2527, Ap
C 41	34	3.0	601	3	US-09-949-016-61404	Sequence 61404, A
C 42	34	3.0	601	3	US-09-949-016-89352	Sequence 89352, A
C 43	34	3.0	714	3	US-09-134-001C-1348	Sequence 1348, Ap
C 44	34	3.0	864	3	US-09-270-767-14160	Sequence 14160, A
C 45	34	3.0	3504	3	US-09-710-279-3859	Sequence 3859, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMUN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PT-F15

US-08-232-463-14

4.9%; Score 55.8; DB 2; Length 7218;

Query Match
Best Local Similarity 3.1%; Pred. No. 4.5e-06;
Matches 12; Conservative 224; Mismatches 151; Indels 0; Gaps 0;

```
OY 197 ATATTATGATGGAGCTATGTTGAAAAAGTAAACGAGCAAAAAGGGCAGACAGTATTT 256
DB 1446 AGAATTTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1387
OY 257 TTCAATATGTTGAACCGATCTTTAGAGCCTCGGATTAATGAGCAGAACTTTGAAAAC 316
DB 1386 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1327
OY 317 CGTAACTATCAAAAAGATTATAAACAAGCAGATAAAGATTATCTGACAGAGATA 376
DB 1326 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1267
OY 377 AGGAATCAGTGAAGTCTTGAAGATATGAATTTCCAGGTTCTCAACAGCCCAACAAC 436
DB 1266 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1207
OY 437 ACGCAATGATTAACGGCTTCAAGGCATGAAAGATACGCTTGAGAGATTGGCAACAA 496
DB 1206 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1147
OY 497 ACCTGATATGCTTGAGCGGATACAGCTTAAGTATGCGCAAAAAGAAATTCGTACC 556
DB 1146 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1087
OY 557 AGAAGTCAACGGGCTTAACGATTGCAA 583
DB 1086 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
```

RESULT 2

US-09-134-001C-1575

Sequence 1575, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NOCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 1575

LENGTH: 1101

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1575

Query Match 4.8%; Score 54.2; DB 3; Length 1101;

Best Local Similarity 4.74%; Pred. No. 4.7e-06;

Matches 229; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

```
OY 465 GAAAGATCGCTTGAGATTGGAGCAAAACCTTGATATCGTTGGAGGGGATAGAC 524
DB 459 GACAGCAATGATGAGAAAATAGATCCAGACGTTATATTTTAAACGGTAATGG 518
OY 525 CTTAAGTATGAGAAAAGAAATTTGTTACAGAAAGTCAACGGGTAAACGATTGCAAC 584
DB 519 TTCAATGCAATTAATAGTAAACCTGTAACAACATTAAGTAAAGTAAAGTAAAGTAA 578
OY 585 GCTTGCTTTACGATGCTGTCGGGAAAGGTTCCGGCTAAAGAAATAGCCGGGGCT 644
DB 579 TGTTCATTTACCGATATGATGATTAAGTAACTTAATCTTTAAAGAACACGACG---TC 635
```

```
OY 645 GCTCCCGCAGATCTGAAATCTTCATCCCTATGATTGAGAACGAAAAACATGCTGA 704
DB 636 AATTAGTTAGATCAGCTATATTTTATCTTTATATAAAAAATTAAAGGAAATTAATGA 695
OY 705 CATGTTGTTGAGTACACTGGGGCAGAGATATGACAAATGATCCAAACGACGCCCA 764
DB 696 TTAGGTGATCTCAATGTAGATGGGGATACCTTAAGAAAGAAATGTGACTACACGTCA 755
OY 765 GCGCAGCTTGACAGAGCATATGTGATGCGGGAGCTGACATCATCTGCGCCATATCC 824
DB 756 AAAAGATATGCAATGCTGTTAGCAATGCTGCGAGATGCTATTATGCTATTAATAC 815
OY 825 GCACGCTTGAACCGATTGAATATATAACGAAACCGTCAATTTCTACAGCCTCGCAA 884
DB 816 --AGTATTAATAAAGTTGAATTAATTAAGCGAACCCATATTTTATAGTTAGGTA 872
OY 885 CTTTGCTTTGACCAAGGCTGAGACGAGAACAGACAGTCACTGTTCACTATCACT 944
DB 873 CACAAGCTGATTAATCTTATCAAAAAATCAAGAAAGAAATGTTGTACAAACACTG 932
OY 945 GAA 947
DB 933 GAA 935
```

RESULT 3

US-09-902-540-6958

Sequence 6958, Application US/09902540

Patent No. 6833447

GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Wiegand, Roger C.

TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(15849)B

CURRENT APPLICATION NUMBER: US/09/902,540

PRIOR FILING DATE: 2001-07-10

PRIOR APPLICATION NUMBER: 60/217,883

PRIOR FILING DATE: 2000-07-10

NUMBER OF SEQ ID NOS: 16825

SEQ ID NO 6958

LENGTH: 939

TYPE: DNA

ORGANISM: Myxococcus xanthus

US-09-902-540-6958

Query Match 4.2%; Score 47.8; DB 3; Length 939;

Best Local Similarity 51.7%; Pred. No. 0.00041;

Matches 109; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
OY 697 CATGTCACATTTGTTGTTGTCACATGACACTGGGGCCCAAGATATGACAAATGATCCAAAC 756
DB 472 CAGCGGACATGCTCTGCGCCCTACTTCCACTGCGGAGCGCAAGCACTACACCCGGAG 531
OY 757 GACCGCAGCGCGAGCTTGCAAGAGCCATGCTGATGCGGGAGCTGACATCATCTGCGAC 816
DB 532 CCTTACCAAGTCCGCTCGGCCCACTGCGCATGCAACGCGGGCGGGGTCTCTGGGC 591
OY 817 CATCATCGCACGCTTATGAACCGATTGAAGTATATAACGAAACCGTCAATTTCTACAGC 876
DB 592 GCCACCGCGACGTCCTCCAGGCGCATGAGCTGTACACAGGCAACGCGGTCTACTCG 651
OY 877 CTGCGCACTTGTCTTTGACCAAGGCTGGA 907
DB 652 CTGGGAACTTCGTCCTTCGGGGGAACCTGGA 682
```

RESULT 4

US-09-902-540-602/c

Sequence 602, Application US/09902540

Patent No. 6833447

GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 602
LENGTH: 4604
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-602

Query Match 4.2%; Score 47.8; DB 3; Length 4604;
Best Local Similarity 51.7%; Pred. No. 0.001;
Matches 109; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

697 CATGTCGACATGTTGTGTGTCAGTCACACTGGGCGCAAGATGACATGATCCAAAC 756
3256 CAGGCGGACATGCTCTCCCTCACTTCCATGGGACCGAAGGACCTACACCCCGAG 3197
757 GACCGCCAGCGCCAGCTTGCAAGACCATGTCTGATCGCGGAGCTGACATCATCGTCGC 816
3196 CCTATACAGTGTCCGCTGGCCACATCGCCATCGACGCGGCGCGGTGTGCTGGGC 3137
817 CATCATCCGACGCTTGTAACCGATGAATATTAAGGAACCGTCAATTTTCTACAGC 876
3136 GCCACCGGACGCTCTCCAGGCGATGAGCTGTACCAAGGACGCGGTGTGCTGTCG 3077
877 CTCGCACTTGTGTTTGACCAAGGCTGA 907
3076 CTGGGAACTGTCTCTTCGGGGGAAGTGA 3046

RESULT 5

US-09-134-001C-1479
Sequence 1479, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1479
LENGTH: 1170
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1479

Query Match 4.2%; Score 47.6; DB 3; Length 1170;
Best Local Similarity 54.6%; Pred. No. 0.00053;
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

686 AACGGAAAAACATGCTGACATGTTGTGTCAGTCACACTGGGCGCAAGATGACGA 745
635 AACGCCAAACAAAGATGATGATCTTTTCAGACATTTGGGATGAAGTAAGGTAAGC 694
746 ATGATCCAAACGACCGCCAGCGCCAGCTTGCAAGAGCAGTGTGATGCGGAGCTGACA 805
695 ATGAGCTTAACGACCTCAACAAAAAATATGCAAAATTTTCGTGATGACAGGTGTGATG 754
806 TCATCGTCGACATCATCGACGCTTTGAACCGATTGAATATTAACGAA 859

755 TAGTTTAGTACACATCCATGATGATTAACCAAGTAATATGGGTTGACCGTA 808

RESULT 6
US-09-710-279-4412
Sequence 4412, Application US/09710279
Patent No. 6703492

GENERAL INFORMATION:
APPLICANT: KIMBERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4412
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4412
LENGTH: 3523
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-4412

Query Match 4.2%; Score 47.6; DB 3; Length 3523;
Best Local Similarity 54.6%; Pred. No. 0.001;
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

686 AACGGAAAAACATGCTGACATGTTGTGTCAGTCACACTGGGCGCAAGATGACGA 745
2635 AACGCCAAACAAAGATGATGATCTTTTCAGACATTTGGGATGAAGTAAGGTAAGC 2694
746 ATGATCCAAACGACCGCCAGCGCCAGCTTGCAAGAGCAGTGTGATGCGGAGCTGACA 805
2695 ATCAGCTTAACGACCTCAACAAAAAATATGCAAAATTTTCGTGATGACAGGTGTGATG 2754
806 TCATCGTCGACATCATCGACGCTTTGAACCGATTGAATATTAACGAA 859
2755 TAGTTTAGTACACATCCATGATGATTAACCAAGTAATATGGGTTGACCGTA 2808

RESULT 7

US-09-214-808-1
Sequence 1, Application US/09214808A
Patent No. 6475793
GENERAL INFORMATION:
APPLICANT: Rosenthal, Andre
APPLICANT: Freiberg, Christoph
APPLICANT: Perret, Xavier Philippe
APPLICANT: Broughton, William John
TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
Patent No. 6475793
FILE REFERENCE: CARP0068
CURRENT APPLICATION NUMBER: US/09/214,808A
CURRENT FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: PCT/IB97/00950
PRIOR FILING DATE: 1997-07-10
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 536165
TYPE: DNA
ORGANISM: Rhizobium
US-09-214-808-1

Query Match 4.1%; Score 47.2; DB 3; Length 536165;
Best Local Similarity 52.0%; Pred. No. 0.024;
Matches 106; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY	705	AATGTGTGTGCGAGTACACTCGGGGCCAAGAGTATGCAATGATATCCAAACGACCGCCA	764
Db	421592	CATTTTTCACAAACACACGACACGAGCCGGGAAATTGAGCGAAGAGCCGGCCATTTTGA	421651
QY	765	GCGCCAGCTTGCAMAGCCATGTCTGATCGGGAGCTGACATCATCTGTGGCCATATCC	824
Db	421652	GCAGGCGCTTGCTCGCAAGCTAATGCATGCGGGAGCGGAGCGCTAGTTGGACACGGACC	421711
QY	825	GCACGCTTTAGAACCGATTGAAGTATATAACGAAACCGTCATTTTCTACAGCTTCGGCAA	884
Db	421712	GCACCGAGTGCCTGCAGATGGAATATCAACAACGCCGGCCCATCTTATAGTCTGGGAA	421771
QY	885	CTTGTGCTTTGACCAAGCTGAGC	908
Db	421772	CTTCTTTATGATGACCTCCGAGC	421795

RESULT 8
US-09-902-540-3551
; Sequence 3551, Application US/09902540

```

1  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
2
3  FILE REFERENCE: 38-10(15849)B
4
5  CURRENT APPLICATION NUMBER: US/09/902,540
6
7  PRIOR FILING DATE: 2001-07-10
8
9  PRIOR APPLICATION NUMBER: 60/217,883
10
11 PRIOR FILING DATE: 2000-07-10
12
13 NUMBER OF SEQ ID NOS: 16825
14
15 SEQ ID NO 3551
16
17 LENGTH: 1116
18
19 TYPE: DNA
20
21 ORGANISM: Myxococcus xanthus
22
23 US-09-902-540-3551

```

Query Match	3.9%;	Score 45;	DB 3;	Length 116;
Best Local Similarity	56.4%;	Pred. No. 0.0033;		
Matches	84;	Conservative	0;	Mismatches 65; Indels

Oy	697	CATGCTGACATTGTTGTTTGTCAGTACAC	CTGGGGGCGAAGATGATGACATGATCCAAAC	75/6
Db	457	CATGTGACGCCCTCATCTGCTGCGGCAC	CTGGGGGACGGAGTACAAAGGGGGAGCCACGG	51/6
Oy	757	GACCGCGACGGCGAGTTTSCAAGAGCCATGTGAT	CGGGGACCTGACATCATCTGTGCGC	81/6
Db	517	CCCGAGAGACCTGTGAGCTGGGACCGCGCGCTCT	GTGGAACGGGGGGCGAAGGGCGGTCAATCGGG	57/6
Oy	817	CATCATCCGCACGCTCTTAGAACCCGATTGA		84/5
Db	577	CACCATCCGCACAGTGTCTCCAGCGCGCTGGA		60/5

RESULT 9
US-09-902-540-1170/c

```

1 Patent NO. 5833447
2 GENERAL INFORMATION:
3 APPLICANT: Goldman, Barry S.
4 APPLICANT: Hinkle, Gregory J.
5 APPLICANT: Slater, Steven C.
6 APPLICANT: Wiegand, Roger C.
7 TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
8 FILE REFERENCE: 38-10(15849)B
9 CURRENT APPLICATION NUMBER: US/09/902,540
10 CURRENT FILING DATE: 2001-07-10
11 PRIOR APPLICATION NUMBER: 60/217, 883
12 PRIOR FILING DATE: 2000-07-10
13 NUMBER OF SEQ ID NOS: 16825
14 SEQ ID NO 1170

```

```

; LENGTH: 17228
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-1170

```

Query Match	3.9%	Score 45;	DB 3;	Length 17228;
Best Local Similarity	56.4%;	Pred. No. 0.016;		
Matches 84;	Conservative	0;	Mismatches 65;	Indels

QY	Db	QY
697	16771	757
CATGTCGACATGTTGTTTGTGACGACACTGCGGGCCAGAAAGTATATGACCATATATCAAC	CATATGACGCGCTCATCTGCTGGCGACCTGGGGAGCGAATTACAGGGGAGGCACGG	GACCCCGACGCGCAGTTGTCAGAAAGCCATGTGTATGTCGGAGCTGACATCATCTGTCGC
756	16712	816
16711	16552	
CCGAGAGACCTGAGTGTGGAGCGCGGCTGTGACAGCGGGGCCAGAGGCGGTATCGGG		

Oy	817	CATCATCCGCACGCTCTAGAACCGATTGA	845
Db	16651	CACCATCCGCACGCTGTCCAGGCCGCTGA	16623

RESULT 10
US-09-949-016-15908/c
; Sequence 15908, Application US/0994901

1. TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

TYPE: DNA
ORGANISM: Human
US-09-949-016-15908

Query Match	3.7%;	Score	42.4;	DB	3;	Length	14554;
Best Local Similarity	50.5%;	Pred. No.	0.092;				
Matches 103; Conservative	0;	Mismatches	101;	Indels	0;	Gaps	0;

QY	368	AGACGATATAGCAATACGTAAAGTCTTGAAGATATGAATTTTCAGGTTCTCAACGCG	427
Db	8850	AAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	8791
QY	428	CCAACAACACGCAATGATTTACGCGCTTCAGGCGATGAAGATACGCTTGAGATTTG	487
Db	8790	AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	8731
QY	488	CGAAGCAAAACCTTGATATCTGTTGGACGGGATACGCTTAAATGATGCCAAAAAGAAA	547
Db	8730	AAAAAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	8671
QY	548	TTTCGTACGAGAAAGTCAACGGGG	571
Db	8670	CTTCCCTTCAGAAAGCCAGCCGAG	8647

```

RESULT 11
US-10-104-047-474
; Sequence 474, Application US/1010404
; Patent No. 6943241
; GENERAL INFORMATION:

```

APPLICANT: HELIX RESEARCH INSTITUTE
 TITLE OF INVENTION: NO. 6943241e1 full length cdna
 FILE REFERENCE: H1-A0105
 CURRENT APPLICATION NUMBER: US/10/104,047
 CURRENT FILING DATE: 2002-03-25
 PRIOR APPLICATION NUMBER:
 PRIOR FILING DATE:
 NUMBER OF SEQ ID NOS: 4096
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 474
 LENGTH: 3256
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-104-047-474

Query Match 3.7%; Score 41.8; DB 3; Length 3256;
 Best Local Similarity 50.2%; Pred. No. 0.059;
 Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 243 GGCAGACGATTTTTCATATGTTGACCCGATCTTTAGAGCTCGGATTATGTACGAG 302
 DB 1684 GGCAATATGCTGCATGATGATCTTGAGACGACCTGAAAGTCTTGTATATGAACGACG 1743
 QY 303 AAACCTTGAACCCGCTAACCTATCAAAAGATTTAAACAAGAGATAAGAGATTCA 362
 DB 1744 TCTGTGTGCTTGGGAGATCTCTGAAACGCTTACCTCTGATGTCAAGGTGC 1803
 QY 363 TCTGACGAGATTAAGATTAAGTGAAGTCTTGAAGATATGATTTTCAAGGTTCTCA 422
 DB 1804 TGACCTGTGAGATTAATCTGTAAGATTTCTTGACGAGATTAATGATATGACCCAG 1863
 QY 423 CAGCGCCAAACACGACGCAATGAT 447
 DB 1864 TGCACGAGGAGAAATAGCAATTGAT 1888

RESULT 12
 US-09-902-540-9266
 ; Sequence 9266, Application US/09902540
 ; Patent No. 6833447
 ; GENERAL INFORMATION:
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15849)B
 ; CURRENT APPLICATION NUMBER: US/09/902,540
 ; CURRENT FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: 60/217,883
 ; PRIOR FILING DATE: 2000-07-10
 ; NUMBER OF SEQ ID NOS: 16825
 ; SEQ ID NO 9266
 ; LENGTH: 1272
 ; TYPE: DNA
 ; ORGANISM: Myxococcus xanthus
 ; US-09-902-540-9266

Query Match 3.5%; Score 39.4; DB 3; Length 1272;
 Best Local Similarity 59.3%; Pred. No. 0.19; 46; Indels 0; Gaps 0;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 779 GAGCCATGCTGTGATGCGGAGCTGACATCATCTGCGCCATCATCCGACGCTTTAGAC 838
 DB 896 GCGCGGTGTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 955
 QY 839 CGATTGAATATATACGAAACCGTCATTTTCTACAGCTTGGCACTTTGTC 891
 DB 956 GCATGAGTTCTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1008

RESULT 13
 US-09-902-540-990

Sequence 990, Application US/09902540
 Patent No. 6833447
 ; GENERAL INFORMATION:
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15849)B
 ; CURRENT APPLICATION NUMBER: US/09/902,540
 ; CURRENT FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: 60/217,883
 ; PRIOR FILING DATE: 2000-07-10
 ; NUMBER OF SEQ ID NOS: 16825
 ; SEQ ID NO 990
 ; LENGTH: 9146
 ; TYPE: DNA
 ; ORGANISM: Myxococcus xanthus
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(9146)
 ; OTHER INFORMATION: unsure at all n locations
 ; US-09-902-540-990

Query Match 3.5%; Score 39.4; DB 3; Length 9146;
 Best Local Similarity 59.3%; Pred. No. 0.59;
 Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
 QY 779 GAGCCATGCTGTGATGCGGAGCTGACATCATCTGCGCCATCATCCGACGCTTTAGAC 838
 DB 6491 GCGCGGTGTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6550
 QY 839 CGATTGAATATATACGAAACCGTCATTTTCTACAGCTTGGCACTTTGTC 891
 DB 6551 GCATGAGTTCTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6603

RESULT 14
 US-09-328-352-1364
 ; Sequence 1364, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 1364
 ; LENGTH: 1785
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-1364

Query Match 3.4%; Score 39; DB 3; Length 1785;
 Best Local Similarity 51.4%; Pred. No. 0.31; 85; Indels 0; Gaps 0;
 Matches 90; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
 QY 375 TAAGATCAGTGAAGTCTTGAAGATGATTTTCAAGGTTCTCAAGCGCCCAACA 434
 DB 1005 TAAATAAATCTTGAAGATTTTGAAGTATCTTAATTTATTTGGTGGCAATATA 1064
 QY 435 CCAAGCATGATTAAGCGGCTTCAAGGCAATGAAGATACGCTTGAAGATTGGCAAGCA 494
 DB 1065 CCAATTTGAAGATTAAGGATGAGCAAGATTAAGCTTATTAATTAACAAGTTGATCAGGC 1124
 QY 495 AAACCTTATATGTTGAGCGGATACAGCTTAAGTATGCAAAAAAAGAAATT 549
 DB 1125 GAGTATTTCTTATATGCGGCTGATGAAACAAAAAGATGTCACAAATTACTTT 1179

RESULT 15


```
US-09-583-110-789/c
; Sequence 789, Application US/09583110
; Patent No. 669703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 789
; LENGTH: 378
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-789

Query Match      3.3%; Score 37.4; DB 3; Length 378;
Best Local Similarity 52.2%; Pred. No. 0.39;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

OY 688 GCGAAAAACATGCTGACATTGTTGTCAGTCACACTGGGCCAAGATGACAAT 747
    |||||
DB 342 GCAGAGAAGAAAGCAGATATCCATATATCATGCTCAGATGGGTGAGATGATTG 283
    |||||

OY 748 GATCGAAGCAGCCGACGCGCAGCTTGCAAGAGCCATGTCGATCGGAGCTGACATC 807
    |||||
DB 282 GAACCACTGAAGAAACAAAAGCTCTTATACAAAGATGATGATTGAGGAGGATATT 223
    |||||

OY 808 ATCGTGGCCATCATCCGACGCTCTTAGAACCGATTGAA 846
    |||||
DB 222 ATCTTGGAGGGCATCTCAGCTTGTGAAACCATCTGAA 184
    |||||
```

Search completed: February 27, 2006, 06:46:19
Job time : 273.458 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 344.29 Seconds
(without alignments) 8289.569 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gacgcagagcagagcagctg.....aagctgcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0, Gapept 1.0

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_hic:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_est7:
9: gb_gsa1:
10: gb_gsa2:
11: gb_gsa3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31.4	51.5	340	8	Z14420	Z14420 CEL13C12 Ch
2	30.8	50.5	334	8	D70055	D70055 CELK095C3F
3	30.8	50.5	360	1	AV193676	AV193676 AV193676
4	30.8	50.5	360	8	D74237	D74237 CELK078EXF
5	30.8	50.5	360	8	D76107	D76107 CELK11402F
6	30.8	50.5	395	8	Z14926	Z14926 CEL9E2 Chr1
7	30.8	50.5	534	3	BJ110239	BJ110239 BJ110239
8	30.8	50.5	534	3	BJ805202	BJ805202 BJ805202
9	30.8	50.5	562	3	BJ763981	BJ763981 BJ763981
10	30.8	50.5	581	3	BJ768324	BJ768324 BJ768324
11	30.8	50.5	583	3	BJ110540	BJ110540 BJ110540
12	30.8	50.5	605	3	BJ117290	BJ117290 BJ117290
13	30.8	50.5	609	1	AU200813	AU200813 AU200813
14	30.8	50.5	610	3	BJ118562	BJ118562 BJ118562
15	30.8	50.5	614	1	AU201470	AU201470 AU201470
16	30.8	49.8	689	10	CW566000	CW566000 OA_Aba009
17	30.4	49.8	817	7	CJ387251	CJ387251 CJ387251
18	30.4	49.8	822	10	CW572172	CW572172 OA_Aba009
19	30.2	49.5	375	7	CN070859	CN070859 1021004D0
20	30.2	49.5	441	5	C27223	C27223 C27223 Rice
21	30.2	49.5	490	7	CV733807	CV733807 FLO--08-O
22	30.2	49.5	493	5	CA009345	CA009345 HD13N21F

23	30.2	49.5	503	7	CV729729	CV729729 FLO--02-O
24	30.2	49.5	528	3	BJ480670	BJ480670 BJ480670
25	30.2	49.5	551	5	BQ461339	BQ461339 HD04J12F
26	30.2	49.5	551	5	BQ658558	BQ658558 HD04J12u
27	30.2	49.5	564	5	CA013187	CA013187 HT07K13F
28	30.2	49.5	581	5	CA010365	CA010365 HT10N17u
29	30.2	49.5	597	2	AJ436212	AJ436212 AJ436212
30	30.2	49.5	637	2	BE194912	BE194912 HVSMEH008
31	30.2	49.5	665	5	BQ768598	BQ768598 EBR08-SQ
32	30.2	49.5	675	7	CR287429	CR287429 CR287429
33	30.2	49.5	677	5	BQ768239	BQ768239 EBR08-SQ
34	30.2	49.5	680	7	CA305627	CA305627 EST000063
35	30.2	49.5	731	7	CV054647	CV054647 BNEB1194
36	30	49.2	514	7	COS37776	COS37776 tah78e07.
37	30	49.2	615	6	CF657343	CF657343 tac74c10.
38	30	49.2	615	8	DN811657	DN811657 ACAC-aab4
39	30	49.2	630	8	DN248864	DN248864 ACAC-aab4
40	30	49.2	728	8	CK634284	CK634284 taf48f11.
41	30	49.2	739	8	DN603980	DN603980 ACAC-aab8
42	29.8	48.9	360	8	D70730	D70730 CEIK123H8F
43	29.6	48.5	544	1	AU056485	AU056485 AU056485
44	29.6	48.5	564	1	AU061922	AU061922 AU061922
45	29.6	48.5	609	1	AU039865	AU039865 AU039865

ALIGNMENTS

RESULT 1
LOCUS Z14420 340 bp mRNA linear EST 19-JUN-1997
DEFINITION CEL13C12 Chris Martin sorted cDNA library Caenorhabditis elegans
CDNA clone cm13c12 5', mRNA sequence.

ACCESSION Z14420.1 GI:6085
VERSION Z14420
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 340)
Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A.,
Hillier, L., Durbin, R.K., Green, P., Showkhen, R., Halloran, N.,
Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J.,
and Sulston, J.
A survey of expressed genes in Caenorhabditis elegans
Nat. Genet. 1, 114-123 (1992)

TITLE A survey of expressed genes in Caenorhabditis elegans
JOURNAL Nat. Genet.
PUBMED 1302004
COMMENT Contact: Waterston R.H. (USA) and Sulston J.E. (UK)
(USA) Dept. of Genetics or (UK) MRC Laboratory of
(USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of
Molecular Biology
Box 8232, 4566 Scott Ave., St. Louis, MI 63110, USA, or, Hills
Road, Cambridge CB2 2QH, UK
Tel: (USA) (314)3627072 or (UK) (0223)248011
Fax: (USA) (314)3624137 or (UK) (0223)402008
Email: twanemalcode.wustl.edu or jse@mc-lmba.cambridge.ac.uk
single read.

FEATURES

source
1..340
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="cm13c12"
/lab_host="MC1061"
/clone_lib="Chris Martin sorted cDNA library"
/note="Vector: lambda phage SLX2. Mixed stage
hermaphrodite cDNA library. Partially normalized by
successively picking groups of clones that didn't
hybridize to previously picked clones. Vector: lambda phage
SLX2 (Lipshitz, D.H. et al., Gene 88:25-36 (1990)) Host:
MC1061"

QY 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGTCTCAAGAACT 58
 |||||
 DB 222 GAGCCAGTGAAGTCCGCTGCTCAAGAGCATCTGACCAAGACGCTGCTCAAGAGCT 279

RESULT 5
 LOCUS D76107 360 bp mRNA linear EST 14-JUL-2005
 DEFINITION CEK114D2P Yui Kohara unpublished cDNA:Strain N2 hermaphrodite embryo *Caenorhabditis elegans* CDNA clone yk114d2 5', mRNA sequence.

ACCESSION D76107
 VERSION D76107.1 GI:1121891
 KEYWORDS EST.
 SOURCE *Caenorhabditis elegans*
 ORGANISM *Caenorhabditis elegans*
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; *Caenorhabditis*.
 1 (bases 1 to 360)
 Kohara,Y., Mitsuki,H., Nishigaki,A., Mochizuki,T., Sugimoto,A. and Tabara,H.
 Toward an expression map of the *C.elegans* genome
 Unpublished (1994)
 Contact: Yui Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

FEATURES
 source
 1..360
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk114d2"
 /sex="hermaphrodite"
 /dev_stage="embryo"
 /clone_lib="Yui Kohara unpublished cDNA:Strain N2 hermaphrodite embryo"

ORIGIN
 Query Match 50.5%; Score 30.8; DB 8; Length 360;
 Best Local Similarity 70.7%; Pred. No. 45;
 Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGTCTCAAGAACT 58
 |||||
 DB 255 GAGCCAGTGAAGTCCGCTGCTCAAGAGCATCTGACCAAGACGCTGCTCAAGAGCT 312

RESULT 6
 LOCUS Z14926 395 bp mRNA linear EST 19-JUN-1997
 DEFINITION CE19E2 Chris Martin sorted cDNA library *Caenorhabditis elegans* CDNA clone cm9e2 5', mRNA sequence.

ACCESSION Z14926
 VERSION Z14926.1 GI:6594
 KEYWORDS EST.
 SOURCE *Caenorhabditis elegans*
 ORGANISM *Caenorhabditis elegans*
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; *Caenorhabditis*.
 1 (bases 1 to 395)
 Waterston,R., Martin,C., Craxton,M., Huynh,C., Coulson,A., Hillier,L., Durbin,R.K., Green,P., Showkneen,R., Hailston,N., Hawkins,T., Wilson,R., Berks,M., Du,Z., Thomas,K., Thierry-Mieg,J. and Sulston,J.
 A survey of expressed genes in *Caenorhabditis elegans*
 Nat. Genet. 1, 114-123 (1992)
 1302004
 Contact: Waterston R.H. (USA) and Sulston J.E. (UK)

(USA) Dept. of Genetics or (UK) (USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of Molecular Biology
 Box 82332, 4566 Scott Ave., St. Louis, MI 63110, USA, or, Hills Road, Cambridge CB2 2QH, UK
 Tel: (USA) (314) 3627072 or (UK) (0223) 248011
 Fax: (USA) (314) 3624137 or (UK) (0223) 402008
 Email: rwenematode.wustl.edu or jesemrc-lmba.cambridge.ac.uk
 single read.

FEATURES
 source
 1..395
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /clone="cm9e2"
 /lab_host="MCI061"
 /clone_lib="Chris Martin sorted cDNA library"
 /note="Vector: lambda phage SHLX2; Mixed stage hermaphrodite cDNA library. Partially normalized by successively picking groups of clones that didn't hybridize to previously picked clones. Vector: lambda phage SHLX2 (Lipschitz, D.H. et al., Gene 88:25-36 (1990)) Host: MCI061"

ORIGIN
 Query Match 50.5%; Score 30.8; DB 8; Length 395;
 Best Local Similarity 70.7%; Pred. No. 46;
 Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGTCTCAAGAACT 58
 |||||
 DB 138 GAGCCAGTGAAGTCCGCTGCTCAAGAGCATCTGACCAAGACGCTGCTCAAGAGCT 195

RESULT 7
 LOCUS BJ110239 534 bp mRNA linear EST 23-JAN-2002
 DEFINITION BJ110239 unpublished oligo-capped cDNA library *C. elegans* L1 stage *Caenorhabditis elegans* CDNA clone yk1128f04 5', mRNA sequence.

ACCESSION BJ110239
 VERSION BJ110239.1 GI:18270276
 KEYWORDS EST.
 SOURCE *Caenorhabditis elegans*
 ORGANISM *Caenorhabditis elegans*
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; *Caenorhabditis*.
 1 (bases 1 to 534)
 Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S.
 A complementary view of the *C.elegans* genome
 Unpublished (2002)
 Contact: Tadaeu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tsnini@genes.nig.ac.jp.

FEATURES
 source
 1..534
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1128f04"
 /sex="hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 /clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"

ORIGIN

Query Match	50.5%	Score 30.8	DB 3	Length 534
Best Local Similarity	70.7%	Fred. No. 48		
Matches	41	Conservative	0	Mismatches 17; Indels 0; Gaps 0;
Oy	1	GATCCAAGTGAAGAAACTGCTCAAGAAACCGCTGCTCAAGAAAGCTGCTCAAGAAACT	58	
Db	265	GAGCCAACTGCAACTCGCTGCTCAAGAAAGCATCTGACCAGAGAGCTGCTGAGAAAGCT	322	
RESULT 8				
BJ805202				
LOCUS	534 bp	mRNA	linear	EST 26-MAY-2004
DEFINITION	BJ805202 unpublished oligo-capped cDNA library, stage I4			
ACCESSION	BJ805202			
VERSION	BJ805202.1	GI:47708869		
KEYWORDS	EST.			
SOURCE	Caenorhabditis elegans			
ORGANISM	Caenorhabditis elegans			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.			
AUTHORS	1 (bases 1 to 534)			
	Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. and Sugano, S.			
TITLE	A complementary view of the C.elegans genome			
JOURNAL	Unpublished (2002)			
CONTACT	Contact: Tadasu Shin-I			
	National Institute of Genetics			
	1111 Yata, Mishima, Shizuoka 411-8540, Japan			
	Tel: 81-559-81-6856			
	Fax: 81-559-81-6855			
	Email: tshin@genes.nig.ac.jp.			
FEATURES	Location/Qualifiers			
source	1..534			
	/organism="Caenorhabditis elegans"			
	/mol_type="mRNA"			
	/strain="N2"			
	/db_xref="taxon:6239"			
	/clone="YK1547a11"			
	/sex="Hermaphrodite"			
	/tissue_type="whole animal"			
	/dev_stage="L4"			
	/clone_idb="unpublished oligo-capped cDNA library, stage L4"			
ORIGIN				
Query Match	50.5%	Score 30.8	DB 3	Length 534
Best Local Similarity	70.7%	Fred. No. 48		
Matches	41	Conservative	0	Mismatches 17; Indels 0; Gaps 0;
Oy	1	GATCCAAGTGAAGAAACTGCTCAAGAAACCGCTGCTCAAGAAAGCTGCTCAAGAAACT	58	
Db	249	GAGCCAACTGCAACTCGCTGCTCAAGAAAGCATCTGACCAGAGAGCTGCTGAGAAAGCT	306	
RESULT 9				
BJ763981				
LOCUS	562 bp	mRNA	linear	EST 25-MAY-2004
DEFINITION	BJ763981 unpublished oligo-capped cDNA library Caenorhabditis			
ACCESSION	BJ763981			
VERSION	BJ763981.1	GI:47624509		
KEYWORDS	EST.			
SOURCE	Caenorhabditis elegans			
ORGANISM	Caenorhabditis elegans			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.			
AUTHORS	1 (bases 1 to 562)			
	Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. and Sugano, S.			
TITLE	A complementary view of the C.elegans genome			
JOURNAL	Unpublished (2002)			

	COMMENT	Contact: Tadasu Shin- Center For Genetic Resource Information National Institute of Genetics . Iii Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. Location/Qualifiers
	FEATURES	1..562 /organism="Caenorhabditis elegans" /mol_type="mRNA" /strain="N2" /db_xref="taxon:6239" /clone="yk1606b07" /bex="hermaphrodite" /tissue_type="whole animal" /dev_stage="varied" /clone_1ib="unpublished oligo-capped cdna library"
	ORIGIN	
Qy	Query Match	50.5%; Score 30.8; DB 3; Length 562; Best Local Similarity 70.7%; Pred. No. 48; Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Db	246	GAGCCAAATGCAAGTGGCTGCTCAAGAACCCTGCTCAAGAGCTGCTCAAGAACT 58 1 GATCACAATGGAAGAAACTGCTCAAGAACCGCTGCTCAAGAGCTGCTCAAGAACT 58 246 GAGCCAAATGCAAGTGGCTGCTCAAGAAGCATCTGACCAAGAGCTGCTCGAAGACT 303
	RESULT 10	
	LOCUS	BUT68324 581 bp mRNA linear EST 25-MAY-2004
	DEFINITION	BUT68324 unpublished oligo-capped cDNA library Caenorhabditis
	ACCESSION	elegans cDNA clone yk1668h11 5', mRNA sequence.
	VERSION	BUT68324
	KEYWORDS	BUT68324.1 GI:47633175
	SOURCE	EST.
	ORGANISM	Caenorhabditis elegans
	REFERENCE	Caenorhabditis elegans Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiidae; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis. 1 (bases 1 to 581) Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S. A complementary view of the C.elegans genome unpublished (2002)
	AUTHORS	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics Iiii Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. Location/Qualifiers
	TITLE	1..581
	JOURNAL	/organism="Caenorhabditis elegans" /mol_type="mRNA" /strain="N2" /db_xref="taxon:6239" /clone="yk1668h11" /bex="hermaphrodite" /tissue_type="whole animal" /dev_stage="varied" /clone_1ib="unpublished oligo-capped cdna library"
	FEATURES	
	source	
	ORIGIN	
Qy	Query Match	50.5%; Score 30.8; DB 3; Length 581; Best Local Similarity 70.7%; Pred. No. 49; Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Db	266	GAGCCAAATGCAAGTGGCTGCTCAAGAAGCATCTGACCAAGAGCTGCTCGAAGACT 323 1 GATCACAATGGAAGAACTGCTCAAGAACCGCTGCTCAAGAGCTGCTCAAGAACT 58 266 GAGCCAAATGCAAGTGGCTGCTCAAGAAGCATCTGACCAAGAGCTGCTCGAAGACT 323

RESULT 11	Bj110540	583 bp	MRNA	linear	EST 23-JAN-2002
LOCUS	Bj110540	unpublished oligo-capped cDNA library, C. elegans l1 stage			
DEFINITION	Bj110540	unpublished oligo-capped cDNA library, C. elegans l1 stage			
ACCESSION	Bj110540	unpublished oligo-capped cDNA library, C. elegans l1 stage			
VERSION	Bj110540.1	GI:18270577			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
FEATURES					
source					
ORIGIN					
Query Match					
Best Local Similarity	50.5%;	Score 30.8;	DB 3;	Length 583;	
Matches	70.7%;	Pred. No. 49;			
Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;	
Cy	1	GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT	58		
Db	266	GAGCCAGTGAAGTGCCTGCTCAAGAAAGCATGTGACCAAGAGCTGCTCAAGAGCT	323		
RESULT 12					
LOCUS	Bj117290	605 bp	MRNA	linear	EST 23-JAN-2002
DEFINITION	Bj117290	unpublished oligo-capped cDNA library, C. elegans l1 stage			
ACCESSION	Bj117290	unpublished oligo-capped cDNA library, C. elegans l1 stage			
VERSION	Bj117290.1	GI:18277403			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					

FEATURES					
source					
Location/Qualifiers					
1..605					
/organism="Caenorhabditis elegans"					
/mol_type="mRNA"					
/strain="N2"					
/db_xref="taxon:6239"					
/clone="yk784c05"					
/sex="hermaphrodite"					
/tissue_type="whole animal"					
/dev_stage="L1"					
/clone_1ib="unpublished oligo-capped cDNA library, C. elegans L1 stage"					
ORIGIN					
Query Match 50.5%; Score 30.8; DB 3; Length 605;					
Best Local Similarity 70.7%; Pred. No. 49;					
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;					
Oy	1	GATCGAAGTGAAGAAGCTGCTCAAGAAACGGTGCTCAAGAGCTGCTCAAGAAACT	58		
Db	266	GAGCCAAGTGCAAGTCGCTGCTCAAGAAAGCATCTGACCAGAGAGCTGCTCAGAGAGCT	323		
RESULT 13					
LOCUS AU200813 609 bp mRNA linear EST 17-JUL-2001					
DEFINITION AU200813 unpublished oligo-capped cDNA library, stage L1					
AU200813 Caenorhabditis elegans cDNA clone YK784c05 5', mRNA sequence.					
ACCESSION AU200813					
VERSION AU200813.1 GI:14829131					
KEYWORDS EST.					
SOURCE Caenorhabditis elegans					
ORGANISM Caenorhabditis elegans					
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea; Rhabditiidae; Pelodierinae; Caenorhabditis. 1 (bases 1 to 609) Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S. A complementary view of the C.elegans genome Unpublished (2001) Contact: Yuji Kohara Genome Biology Lab. National Institute of Genetics Yata I111, Mishima, Shizuoka 411, Japan Tel: 81-559-81-6854 Fax: 81-559-81-6855 Email: ykohara@lab.nig.ac.jp. location/Qualifiers					
FEATURES					
source					
1..609					
/organism="Caenorhabditis elegans"					
/mol_type="mRNA"					
/strain="N2"					
/db_xref="taxon:6239"					
/clone="yk784c05"					
/sex="Hermaphrodite"					
/tissue_type="Whole animal"					
/dev_stage="L1"					
/clone_1ib="unpublished oligo-capped cDNA library, stage L1"					
ORIGIN					
Query Match 50.5%; Score 30.8; DB 1; Length 609;					
Best Local Similarity 70.7%; Pred. No. 49;					
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;					
Oy	1	GATCGAAGTGAAGAAGCTGCTCAAGAAACGGTGCTCAAGAGCTGCTCAAGAAACT	58		
Db	266	GAGCCAAGTGCAAGTCGCTGCTCAAGAAAGCATCTGACCAGAGAGCTGCTCAGAGAGCT	323		
RESULT 14					
LOCUS BJI18562 610 bp mRNA linear EST 23-JAN-2002					

```

/db xref="taxon:6239"
/clone="YK793e08"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_idb="unpublished oligo-capped cdna library, stage
L1"

ORIGIN

Query Match          50.5%; Score 30.8; DB 1; Length 614;
Best Local Similarity 70.7%; Pred. No. 49;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Oy      1  GATCCAAGTGAAGAACTGCTCAAGAACCGTGTCTCAAGAGCTGCTCAAGAACT 58
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      264 GAGCCAAGTCAACAGCTGCTCTCAAGAACATCTGACCAAGAGAGCTGTGAGAAAGCT 321

Search completed: February 27, 2006, 11:09:33
Job time : 348.29 secs

```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocellacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time: 38.5097 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61

Sequence: 1 gatccaagtgaagaactg.....aagctgctcgaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004as:*
- 13: geneseqn2004bs:*
- 14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	61	100.0	61	13	ADV65735	Adv65735 Peptide p
2	46	75.4	62	13	ADV65736	Adv65736 Oligonuc
3	30.4	49.8	882	12	AD130094	Ad130094 M. methyl
4	30.2	49.5	549	8	ADAT0869	Ada70869 Rice gene
5	29.4	48.2	773	3	ACAC3118	Acac3118 Arabidops
6	29.4	48.2	787	6	ABN99086	Abn99086 Arabidops
7	29.4	48.2	827	3	ACAC39382	Acac39382 Arabidops
8	29.4	48.2	872	3	ACAC51288	Acac51288 Arabidops
9	29.4	48.2	874	3	ACAC37192	Acac37192 Arabidops
10	29.4	48.2	1420	3	AAW78510	Aaw78510 Plant SDR
11	29.2	47.9	1372	13	ADS50247	Ads50247 Bacterial
12	28.6	46.9	296	6	ABL70950	Ab170950 Corn case
13	28.6	46.9	546	11	ACU33808	Ac133808 Rice abio
14	28.6	46.9	546	12	ADU44427	Adj44427 Plant CDN
15	28.6	46.9	785	11	ACU33278	Ac133278 Rice abio
16	28.6	46.9	785	12	ADU43613	Adj43613 Plant CDN
17	28.6	46.9	866	13	ADX46590	Adx46590 Plant full
18	28.6	46.9	1019	12	ADJ43614	Adj43614 Plant CDN
19	28.2	46.2	1592	2	AAQ14501	Aaq14501 USP45 gen

20	28.2	46.2	1920	2	AAQ13830	Aaq13830 MSP gene
21	28	45.9	916	13	ADS47519	Ads47519 Bacterial
22	27.8	45.6	2103	10	ADC91813	Adc91813 E. faeciu
23	27.8	45.6	160755	4	AAH88704	Aah88704 Human DNA
24	27.6	45.2	3870	4	ABL28047	Ab128047 Drosophi
25	27.6	45.2	6556	4	ABL28046	Ab128046 Drosophi
26	27.4	44.9	469	9	ACH14859	Ach14859 Human adu
27	27.4	44.9	1227	4	AA160808	Aa160808 Human pol
28	27.4	44.9	1466	4	AA159022	Aa159022 Human pol
29	27.4	44.9	1466	5	ADO99244	Ado99244 DNA encod
30	27.4	44.9	1466	9	ADB49004	Adb49004 Novel hum
31	27.4	44.9	1473	12	ADP28487	Adp28487 Human sec
32	27.4	44.9	2078	13	ACN43611	Acn43611 Human dia
33	27.4	44.9	2089	13	ACN43610	Acn43610 Human dia
34	27.4	44.9	2093	13	ACN43609	Acn43609 Human dia
35	27.4	44.9	2155	13	ACN43608	Acn43608 Human dia
36	27.4	44.9	2203	13	ACN43607	Acn43607 Human dia
37	27.4	44.9	2207	13	ACN43606	Acn43606 Human dia
38	27.4	44.9	2216	13	ACN43605	Acn43605 Human dia
39	27.4	44.9	2240	13	ACN43604	Acn43604 Human dia
40	27.4	44.9	2241	13	ACN43603	Acn43603 Human dia
41	27.4	44.9	2245	13	ACN43602	Acn43602 Human dia
42	27.4	44.9	2256	12	ADJ27218	Adj27218 Human TRI
43	27.4	44.9	2272	13	ACN43614	Acn43614 Human dia
44	27.4	44.9	2282	13	ACN43613	Acn43613 Human dia
45	27.4	44.9	2300	13	ACN43612	Acn43612 Human dia

ALIGNMENTS

RESULT 1	ADV65735	ADV65735 standard; DNA; 61 BP.
ID	ADV65735	
XX	ADV65735;	
AC	10-FEB-2005 (first entry)	
DT	Peptide p5 encoding DNA SEQ ID NO:4.	
XX	de; gene expression; antibiotic; antimicrobial; fungicide; cytostatic;	
KW	peptide p5.	
OS	Synthetic.	
XX	KR2004034780-A.	
PN	29-APR-2004.	
XX	17-OCT-2002; 2002KR-00063379.	
PF	17-OCT-2002; 2002KR-00063379.	
XX	17-OCT-2002; 2002KR-00063379.	
PR	(BIOL-) BIOLEADERS CORP.	
PA	(UYCH-) UNIV CHUNGSUN CO. LTD.	
PA	(KORE-) KOREA RES INST BIOTECNOLOG.	
XX	Boo HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;	
PI	Park YG, Sung NH;	
XX	WPI, 2004-577380/56.	
DR	Method for surface expression of peptides p5 and anal3 using pgs bca	
XX	gene.	
PT	Claim 3; SEQ ID NO 4; 25pp; Korean.	
PS	The invention relates to a novel method for surface expression of	
XX	peptides p5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs)BCA	
CC	gene, thereby removing a purification process of peptides p5 and Anal3,	
CC	and using lactic acid bacteria for the surface expression, so that	
CC	peptide antibiotics can be cheaply and stably mass-produced. An	

CC expression vector pHCeILB:pgsa-P5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide P5 encoded by the nucleotide sequence set forth
CC in ADVE5735, or to the peptide Ana13 encoded by the nucleotide sequence
CC set forth in ADVE5737. The present sequence encodes the peptide p5 used
CC in the invention.

XX Sequence 61 BP; 23 A; 14 C; 14 G; 10 T; 0 U; 0 Other;
SQ

Query Match 100.0%; Score 61; DB 13; Length 61;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATCGAATGGAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 60
DB 1 GATCGAATGGAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 60
OY 61 A 61
DB 61 A 61

RESULT 2
ADVE5736/C
ID ADV65736 standard; DNA; 62 BP.
AC ADV65736;
XX
DT 10-FEB-2005 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:5.
XX
KM ss; gene expression; antibiotic; antimicrobial; fungicide; cytostatic.
XX
OS Synthetic.
XX
PM KR2004034780-A.
PD 29-APR-2004.
XX
PF 17-OCT-2002; 2002KR-00063379.
XX
PR 17-OCT-2002; 2002KR-00063379.
XX
PA (BIOL-) BIOLEADERS CORP.
PA (UYCH-) UNIV CHUNSUN CO LTD.
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
PI BOC HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
PI Park YG, Sung MH;
XX
DR WPI; 2004-577380/56.
XX
PT Method for surface expression of peptides p5 and ana13 using pgs bca
XX gene.
XX
PS Example 1; SEQ ID NO 5; 25pp; Korean.
XX
CC The invention relates to a novel method for surface expression of
CC peptides p5 and Ana13 using a poly-gamma-glutamic acid synthase (pgs)BCA
CC gene, thereby removing a purification process of peptides p5 and Ana13,
CC and using lactic acid bacteria for the surface expression, so that
CC peptide antibiotics can be cheaply and stably mass-produced. An
CC expression vector pHCeILB:pgsa-P5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide P5 encoded by the nucleotide sequence set forth
CC in ADVE5735, or to the peptide Ana13 encoded by the nucleotide sequence
CC set forth in ADVE5737. The present sequence represents an oligonucleotide
CC used in the invention to introduce the peptide p5 encoding DNA into

CC expression vector pHCeILB:A-P5.
XX
SQ Sequence 62 BP; 11 A; 15 C; 14 G; 22 T; 0 U; 0 Other;
SQ

Query Match 75.4%; Score 46; DB 13; Length 62;
Best Local Similarity 98.3%; Pred. No. 2.3e-05;
Matches 57; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 5 CAGTGGAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 61
DB 62 CAGTGGAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 5

RESULT 3
AD130094
ID AD130094 standard; DNA; 882 BP.
AC AD130094;
XX
DT 22-APR-2004 (first entry)
XX
DE M. methylotrophus D-arabino-3-hexulose 6-phosphate synthase (hps1B) DNA.
XX
KM C1 assimilation; L-amino acid production; fermentation; gene; ds.
XX
OS Methylotrophus methylotrophus.
XX
FH Key Location/Qualifiers
FT CDS 1..882
FT /*tag= a
FT /product= "M. methylotrophus enzyme involved in C1
FT assimilation"
XX
PM US200323238-A1.
XX
PD 18-DEC-2003.
XX
XX 12-JUN-2002; 2002US-0016653.
XX
PF 12-JUN-2002; 2002US-0016653.
XX
PR 12-JUN-2002; 2002US-0016653.
XX
XX (USUD/) USUDA Y.
XX (NISH/) NISHIO Y.
XX (YASU/) YASUEDA H.
XX (SUGI/) SUGIMOTO S.
XX
PI Usuda Y, Nishio Y, Yasueda H, Sugimoto S;
XX
DR WPI; 2004-061268/06.
XX
DR P-PSDB; AD130095.
XX
PT New polynucleotides encoding polypeptides involved in one-carbon compound
PT metabolism in Methylotrophus methylotrophus, useful for producing L-amino
XX acids (e.g. threonine or serine) or as probes and/or as primers.
XX
PS Claim 13; SEQ ID NO 11; 53pp; English.
XX
XX The invention relates to enzymes involved in C1 assimilation in
CC Methylotrophus methylotrophus and their corresponding polynucleotides. The
CC invention also relates to methods of producing amino acids in micro-
CC organisms having enhanced or attenuated expression of the polypeptides
CC and polynucleotides involved in C1 assimilation. The invention is useful
CC for producing L-amino acids (e.g. asparagine, threonine or serine) using
CC Methylotrophus bacteria. It is also used in identifying microorganisms
CC that can be used to produce fine chemicals, for e.g. by fermentative
CC processes. The present sequence is M. methylotrophus gene encoding an
XX enzyme involved in C1 assimilation.
SQ

Query Match 49.8%; Score 30.4; DB 12; Length 882;
Best Local Similarity 71.4%; Pred. No. 4.6;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 6 AAGTGGAGAACTGCTCAAGAAACCCCTGCTCAAGAACTGCTCAAGAACTGTA 61
DB 815 AAGAGGCTGCACTGCTGTAAGCTGCGCCCTGTAAGAACTGCGCGACGAACTGAA 870

RESULT 4

ADA70869
ID ADA70869 standard; DNA; 549 BP.

ADA70869;

20-NOV-2003 (first entry)

Rice gene, SEQ ID 4192.

Plant; bacterial infection; fungal infection; viral infection; rice;
gene; ds.

Oryza sativa.

WO2003000898-A1.

03-JAN-2003.

22-JUN-2001; 2001WO-IB001105.

22-JUN-2001; 2001WO-IB001105.

(SYGN) SYNGENTA PARTICIPATIONS AG.

Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

WPI; 2003-175290/17.

Identifying at least one gene involved in plant resistance or response to
pathogenic infection for conferring resistance or tolerance to a plant to
bacterial, fungal or viral infection by determining or detecting plant
gene expression.

Claim 6; SEQ ID NO 4192; 899bp; English.

The present invention relates to a method (M1) for identifying genes
involved in plant resistance or response to pathogenic infection. M1
comprises identifying a gene whose expression is significantly altered in
the incompatible interaction of plant gene expression relative to
expression of the gene in an uninfected plant, in a mutant plant that
does not express a gene associated with response to pathogenic infection,
or in a corresponding incompatible or compatible interaction. (M1) is
useful for conferring resistance to resistance or tolerance to a plant to
bacterial, fungal or viral infection. The present sequence was used to
illustrate the invention.

Sequence 549 BP; 141 A; 122 C; 171 G; 115 T; 0 U; 0 Other;

Query Match 49.5%; Score 30.2; DB 8; Length 549;

Best Local Similarity 74.5%; Pred. No. 4.8;

Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 11 GAAGAACTGCTCAAGAAACCCCTGCTCAAGAACTGCTCAAGAACTGTA 61

DB 205 GATGAACCTGTGAAGAGCGAGCTGCTGTAAGCTATTGAAGAACTGGA 255

RESULT 5

AAC35118
ID AAC35118 standard; DNA; 773 BP.

AAC35118;

17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9060.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.

OS Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.

05-MAR-1999; 99US-0123180P.

09-MAR-1999; 99US-0123548P.

23-MAR-1999; 99US-0125788P.

25-MAR-1999; 99US-0126264P.

29-MAR-1999; 99US-0126785P.

01-APR-1999; 99US-0127462P.

06-APR-1999; 99US-0128234P.

08-APR-1999; 99US-0128714P.

16-APR-1999; 99US-0129845P.

19-APR-1999; 99US-0130077P.

21-APR-1999; 99US-0130449P.

23-APR-1999; 99US-0130510P.

28-APR-1999; 99US-0130891P.

30-APR-1999; 99US-0131449P.

04-MAY-1999; 99US-0132407P.

05-MAY-1999; 99US-0132484P.

06-MAY-1999; 99US-0132486P.

07-MAY-1999; 99US-0132863P.

11-MAY-1999; 99US-0134256P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134219P.

18-MAY-1999; 99US-0134706P.

19-MAY-1999; 99US-0134941P.

20-MAY-1999; 99US-0135124P.

21-MAY-1999; 99US-0135133P.

24-MAY-1999; 99US-0135629P.

25-MAY-1999; 99US-0136021P.

27-MAY-1999; 99US-0136392P.

28-MAY-1999; 99US-0136782P.

01-JUN-1999; 99US-0137222P.

03-JUN-1999; 99US-0137528P.

04-JUN-1999; 99US-0137502P.

07-JUN-1999; 99US-0137724P.

08-JUN-1999; 99US-0138094P.

10-JUN-1999; 99US-0138540P.

10-JUN-1999; 99US-0138847P.

14-JUN-1999; 99US-0139119P.

16-JUN-1999; 99US-0139452P.

16-JUN-1999; 99US-0139453P.

17-JUN-1999; 99US-0139492P.

18-JUN-1999; 99US-0139454P.

18-JUN-1999; 99US-0139455P.

18-JUN-1999; 99US-0139456P.

18-JUN-1999; 99US-0139457P.

18-JUN-1999; 99US-0139458P.

18-JUN-1999; 99US-0139459P.

18-JUN-1999; 99US-0139460P.

18-JUN-1999; 99US-0139461P.

18-JUN-1999; 99US-0139462P.

18-JUN-1999; 99US-0139463P.

18-JUN-1999; 99US-0139750P.

18-JUN-1999; 99US-0139763P.

21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144336P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.

PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152353P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156599P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 46.2%; Score 29.4; DB 3; Length 773;

Best Local Similarity 76.6%; Pred. No. 9.4; Mismatches 11; Indels 0; Gaps 0;

Matches 36; Conservative 0;

Qy 11 GAAGAACTGCTCAAGAAACGCTGCTCAAGAACTGCTCAAGAAAC 57
Db 424 GAAGAAATATATGACGAAACGCTAGTGAAGAACTTGTGAAGAAAC 470

RESULT 6

ID AEN99086
ABN99086 standard; DNA; 787 BP.

AC AEN99086;

DT 01-AUG-2002 (first entry)

DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 854.

KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
disease; crop; thale cress; tolerance factor; insect; pathogen;

XX nutrition; ds.

OS Arabidopsis thaliana.

XX US2002023281-A1.
XX
XX 21-FEB-2002.
XX
XX 26-JAN-2001; 2001US-00770445.
XX
XX 27-JAN-2000; 2000US-0178472P.
XX
XX (GORLACH J.
XX (ANYU/ AN Y.
XX (HAMILTON C M.
XX (PRICE J L.
XX (RAINE/ RAINES T M.
XX (YUYU/ YU Y.
XX (RAMEKA J G.
XX (PAGE/ PAGE A.
XX (MATH/ MATHW A V.
XX (LEDF/ LEDFORD B L.
XX (WOES/ WOESSNER J P.
XX (HAAS/ HAAS W D.
XX (GARC/ GARCIA C A.
XX (KRICK/ KRICKER M.
XX (SLAT/ SLATER T.
XX (DAVI/ DAVIS K R.
XX (ALLE/ ALLEN K.
XX (HOFF/ HOFFMAN N.
XX (HURB/ HURBAN P.
XX
XX Goriach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
XX Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
XX Hurban P;
XX
XX MPI; 2002-403163/43.
XX
XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,
XX producing compositions that modulate the expression or function of its
XX encoded protein, and mapping functional regions of protein.
XX
XX Claim 1; SEQ ID NO 854; 49pp + Sequence Listing; English.
XX
XX The invention relates to an Arabidopsis thaliana nucleic acid (I)
XX comprising a sequence capable of hybridizing under stringent conditions
XX to a sequence selected from any one of 999 sequences (ABN98233-ABN99233),
XX given in the specification or its fragment. A polypeptide (II) encoded by
XX (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
XX genetically modified cell (IV) comprising an exogenous nucleic acid, is
XX useful for screening a candidate agent for its biological effect. (I) is
XX useful in identifying homologous or related genes, in producing
XX protein, mapping functional regions of the protein and in studying
XX associated physiological pathways. (I) is also useful for the genetic
XX manipulation of cells, particularly plant cells. (I) is also useful in
XX screening assays of various plant strains to determine the strains that
XX are best capable of withstanding a particular disease or environmental
XX stress. (II) and (III) are useful for screening of biologically active
XX agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
XX pathways. The screened agents are useful in improved methods of treating
XX crops to prevent or treat disease. (II) are also useful in screening
XX programs to identify agents that mimic or enhance the action of tolerance
XX factors. Such agents are useful in improved methods of treating crops to
XX enhance their tolerance to environmental stress. (I) is also useful for
XX enhancing or inhibiting production of a biosynthetic product in a plant.
XX (III) is useful for identifying other mediators that may induce
XX expression of proteins of interest, for establishing the extent to which
XX any specific insect and/or pathogen is responsible for damage to a
XX particular plant, for identifying other mediators that enhance or induce
XX tolerance to environmental stress, for identifying factors involved in
XX biosynthetic pathways of nutritional, commercial, or medicinal value and
XX for identifying productions of nutritional, commercial or medicinal
XX value. (IV) is useful in the study of genetic function and regulation,
XX for alteration of the cellular metabolism and for screening compounds

CC that may affect the biological function of the gene or gene products.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=999909770445
XX
XX SQ Sequence 787 BP; 241 A; 156 C; 173 G; 217 T; 0 U; 0 Other;
XX
XX Query Match 48.2%; Score 29.4; DB 6; Length 787;
XX Best Local Similarity 76.6%; Pred. No. 9.4;
XX Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
XX QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAAAC 57
XX DB 280 GAAGAAATATATGACGAAACCGCTAGTGAAGAAAGTTGCTGAAGAAAC 326
XX
XX RESULT 7
XX AAC39382 standard; DNA; 827 BP.
XX
XX AAC39382;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 24429.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132613P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135132P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137727P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-014154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145275P.
PR 27-JUL-1999; 99US-0145911P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147955P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148555P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149729P.
PR 20-AUG-1999; 99US-0149922P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150864P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151067P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153788P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-015659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 48.2%; Score 29.4; DB 3; Length 827;
Best Local Similarity 76.6%; Pred. No. 9.5;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

11
363 GAGGAATATATGACCAACCGTACTGATGAAAGTTGCTGAGGAAC 409

RESULT 8
AAC51288
ID AAC51288 standard; DNA; 872 BP.
XX
AC AAC51288;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67963.
XX
XX Hybridization assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129645P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0141768P.
PR 19-MAY-1999; 99US-0143941P.
PR 20-MAY-1999; 99US-0145124P.
PR 21-MAY-1999; 99US-0145353P.
PR 24-MAY-1999; 99US-0145629P.
PR 25-MAY-1999; 99US-0146021P.
PR 27-MAY-1999; 99US-0146392P.
PR 28-MAY-1999; 99US-0146782P.
PR 01-JUN-1999; 99US-0147222P.
PR 03-JUN-1999; 99US-0147528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 18-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 22-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144844P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 23-JUL-1999; 99US-0145276P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 02-AUG-1999; 99US-0146389P.
PR 02-AUG-1999; 99US-0146389P.
PR 02-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 05-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.


```
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154039P.
PR 16-SEP-1999; 99US-0154779P.
PR 20-SEP-1999; 99US-0155139P.
PR 22-SEP-1999; 99US-0155486P.
PR 23-SEP-1999; 99US-0155659P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159328P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 14-OCT-1999; 99US-0159584P.
PR 18-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
```

Query Match 48.2%; Score 29.4; DB 3; Length 872;
Best Local Similarity 76.6%; Pred. No. 9.7;

Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
OY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAAAC 57
.DB 364 GAAGAAATTAATGACGAAACCGCTAGTGAAGATTGCTGAAGAAAC 410
```

```
RESULT 9
AAC37192
ID AAC37192 standard; DNA; 874 BP.
XX
AC AAC37192;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 16506.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130549P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 03-JUN-1999; 99US-0137528P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
```

PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 24-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-014154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 09-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144884P.
 PR 21-JUL-1999; 99US-0144814P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 26-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147955P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153768P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0156559P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 22-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 25-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161932P.
 PR 28-OCT-1999; 99US-0161933P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 48.2%; Score 29.4; DB 3; Length 874;
 Best Local Similarity 76.6%; Pred. No. 9.7;
 Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 11 GAGAACTGCTCAAGAAACCGCTGCTCAAGAAAGCTGCTCAAGAAC 57
 Db 366 GAGAAATATATACAGAAACGCTAGTGAAGAAAGTGTGGAAGAAC 412

RESULT 10
 ID AAA78510 standard; DNA; 1420 BP.
 XX
 AC AAA78510;

XX 06-AUG-2003 (revised)
DT 27-NOV-2000 (first entry)
XX
DE Plant SDF polynucleotide sequence SEQ List 1 NO:272.
XX
XX Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment; SDF;
KW genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control; ds.
XX
OS Viridiplantae.
XX
XX MO200040695-A2.
XX
XX 13-JUL-2000.
XX
XX 07-JAN-2000; 2000MO-US000466.
XX
XX 08-JAN-1999; 99US-0115293P.
XX
XX (CERE-) CERES INC.
XX
XX Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;
XX
XX MPI; 2000-465970/40.
XX
XX New corn plant and Arabidopsis thaliana sequence-determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene.
XX
XX Claim 1; Page 453-454; 673pp; English.
XX
XX The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to as
CC sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a target
CC gene, either as a promoter, a structural gene, an UTR or as a 3'
CC termination sequence. They are also useful as tools for genetic mapping,
CC and identification of a particular individual plant or for clustering a
CC group of plants with a common trait. AAAT78433 to AAAT78630 and AAB24605 to
CC AAB25009 represent the specifically claimed polynucleotide sequences and
CC polypeptides encoded by them given in the present invention. (Updated on
CC 06-AUG-2003 to correct OS field.)
XX
XX Sequence 1420 BP; 377 A; 287 C; 348 G; 408 T; 0 U; 0 Other;
SQ
Query Match 48.2%; Score 29.4; DB 3; Length 1420;
Best Local Similarity 76.6%; Pred. No. 11;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 11 GAGAAACTGCTCAAGAAACCGCTGCTCAGAGAGCTGCTCAAGAAAC 57
DB 359 GAAAGAAATATATGACGAAACCGCTGTAAGAGATTGCTGAAGAAAC 405
RESULT 11
ID ADS50247 standard; cDNA; 1372 BP.
XX
XX ADS50247;
AC
DT 02-DEC-2004 (first entry)
XX
XX Bacterial polynucleotide #4990.
XX
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
XX Bacteria.
XX
XX US2003233675-A1.
XX
XX 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
XX
XX 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
XX
XX (HINKLE/) HINKLE G J.
XX
XX (SLATER/) SLATER S C.
XX
XX (CHEN/) CHEN X.
XX
XX (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI
XX
XX MPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 28677; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1372 BP; 393 A; 312 C; 322 G; 345 T; 0 U; 0 Other;
SQ
Query Match 47.9%; Score 29.2; DB 13; Length 1372;
Best Local Similarity 69.0%; Pred. No. 12;
Matches 40; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 GATCCAGTGAAGAAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58
DB 547 GAGCCAACTGCAATCCGCTGCTCAAGAAAGCATTTGACCAAGAGAGTGTGAGAAAGCT 604
RESULT 12
ID ABL70950 standard; cDNA; 296 BP.
XX
XX ABL70950;
AC
XX
DT 14-MAY-2002 (first entry)

XX DE Corn tassal-derived polynucleotide (cdps) SEQ ID NO:324.
 XX XX
 XX XX Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPs;
 XX XX inheritance; characteristic; growth; development; disease resistance;
 XX XX environmental adaptability; quality; yield; molecular marker;
 XX XX multigene trait; plant breeding; corn tassal; gene; ss.
 XX XX
 OS Zea mays.
 XX XX
 XX PN US2001051335-A1.
 XX XX
 PD 13-DEC-2001.
 XX XX
 XX PF 16-APR-1999; 99US-00294093.
 XX XX
 XX PR 21-APR-1998; 98US-0082567P.
 XX XX
 XX PA (LAIG/) LAIGUDI R V.
 XX PA (ITOL/) ITO L Y.
 XX PA (SHER/) SHERMAN B K.
 XX PI
 XX PI Lalgudi RV, Ito LY, Sherman BK;
 XX XX
 DR WPI, 2002-163647/21.
 XX XX
 XX PT Novel purified corn tassal-derived polynucleotide useful for determining
 PT altered gene expression, to recover regulatory elements and to follow
 PT inheritance of desirable characteristics through hybrid breeding
 PT programs.
 XX XX
 PS Claim 1; SEQ ID NO 324; 201pp; English.
 XX XX
 CC The present sequence describes a purified corn tassal-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABU70627 to ABU76833. The cdps sequences
 CC can be used for determining altered gene expression, to recover
 CC encode corn tassal-derived polypeptides (CDPs). The cdps sequences (1)
 CC can be used for determining altered gene expression, to recover
 CC characteristic elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (1) are also useful in
 CC the evaluation, and alteration of desired characteristics associated with
 CC growth and development, disease resistance, environmental adaptability
 CC quality and yield, and as molecular markers for studying inheritance of
 CC multigene traits in a plant breeding program. (1) can be used to produce
 CC a tassal-specific profile of gene transcription, a transcript image, to
 CC clone regulatory elements for use in transformation vectors, to express a
 CC polypeptide, to identify, isolate or extend identical or related corn
 CC tassal nucleic acid sequences from DNA libraries, in nucleic acid
 CC hybridisation or amplification technologies, as query sequences to
 CC determine homology of known sequences, as probe for use in Southern or
 CC Northern hybridisation, and to identify the presence of and/or to
 CC determine the degree of similarity between two (or more) nucleic acid
 CC sequences
 XX XX
 SQ Sequence 296 BP; 77 A; 62 C; 97 G; 60 T; 0 U; 0 Other;
 XX XX
 XX Query Match 46.9%; Score 28.6; DB 6; Length 296;
 XX Best Local Similarity 72.5%; Pred. No. 14;
 XX Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 11 GAAGAACTGCTCAAGAACCGCTGCTCAAGAAAGCTGCTCAAGAACTGTA 61
 DB 16 GATGAAACTGTGAGAGCGACGAGCTCGAGAAAGCTATAGAGAAAGCTGGA 66
 RESULT 13
 AC133808
 ID AC133808 standard; cDNA; 546 BP.
 XX
 AC AC133808;
 XX
 DT 02-JUN-2005 (first entry)
 XX

DE DE Rice abiotic stress response related polynucleotide SEQ ID NO:12371.
 XX XX
 XX XX ss; abiotic stress tolerance; transgenic plant; plant; cereal;
 XX XX agriculture.
 XX XX
 OS Oryza sativa.
 XX XX
 XX PN WO2003008540-A2.
 XX XX
 XX PD 30-JAN-2003.
 XX XX
 XX PF 21-JUN-2002; 2002WO-US019668.
 XX XX
 XX PR 22-JUN-2001; 2001US-0300112P.
 XX PR 24-AUG-2001; 2001US-0314662P.
 XX PR 26-SEP-2001; 2001US-0325277P.
 XX PR 21-NOV-2001; 2001US-0332132P.
 XX XX
 XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX XX
 XX PI Krepes J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
 XX PI Moughamer T, Provart N, Rieke D, Zhu T;
 XX XX
 DR WPI, 2003-248011/24.
 XX XX
 XX PT New stress-responsive nucleic acid, useful for altering the
 PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
 PT stress, salt stress or osmotic stress.
 XX XX
 PS Claim 48; SEQ ID NO 12371; 89pp; English.
 XX XX
 CC The invention relates to novel abiotic stress responsive polynucleotides
 CC and polypeptides. Also disclosed are vectors, expression cassettes, host
 CC cells, and plants containing such polynucleotides. Also disclosed are
 CC methods for using the polynucleotides and polypeptides to alter the
 CC responsiveness of a plant to abiotic stress. The invention is useful in
 CC agriculture. The nucleic acid is useful for determining whether a test
 CC plant has been exposed to an abiotic stress condition. It is also useful
 CC for selecting an agent that alters abiotic stress regulated
 CC polynucleotide expression in a plant cell, and to identify a homolog or
 CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
 CC molecule and the polypeptide encoded by it are useful in altering the
 CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
 CC stress, osmotic stress or any of their combinations. The present sequence
 CC is used in the exemplification of the invention
 XX XX
 SQ Sequence 546 BP; 143 A; 123 C; 169 G; 111 T; 0 U; 0 Other;
 XX XX
 XX Query Match 46.9%; Score 28.6; DB 11; Length 546;
 XX Best Local Similarity 72.5%; Pred. No. 16;
 XX Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAAAGCTGCTCAAGAACTGTA 61
 DB 205 GATGAAACTGTGAGAGCGACGAGCTCGAGAAAGCTATAGAGAAAGCTGGA 255
 RESULT 14
 ADJ44427
 ID ADJ44427 standard; cDNA; 546 BP.
 XX
 AC ADJ44427;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 XX Plant cDNA #5427.
 XX
 XX Plant; gene; ss; transcription; plant genome augmentation; cereal;
 XX soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
 XX maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
 XX stress tolerance; salt tolerance; cold tolerance; drought tolerance;
 XX plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
 XX antifungal.

```

XX OS Eukaryota.
XX PN US2004016025-A1.
XX PD 22-JAN-2004.
XX PF 26-SEP-2002; 2002US-00260238.
XX PR 26-SEP-2001; 2001US-0325277P.
XX PR 26-SEP-2001; 2001US-0325448P.
XX PR 04-APR-2002; 2002US-0370620P.
XX PA (BUDWORTH P.
XX PA (MOUGHAMER T.
XX PA (BRIGGS S P.
XX PA (COOPER B.
XX PA (GLAZEBROOK J.
XX PA (GOFF S A.
XX PA (KATAGIRI F.
XX PA (KREPS J.
XX PA (PROVART N.
XX PA (RICE D.
XX PA (ZHUTU) ZHU T.
XX PI Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
XX PI Goff SA, Katagiri F, Kreps J, Provart N, Riche D, Zhu T;
XX DR WPI; 2004-190374/18.
XX PT New rice promoter, useful for manipulating crop plants to alter or
XX PT improve phenotypic characteristics, e.g. produce large quantities of oil
XX PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
XX PT or high nutritional value.
XX PS Example 13; SEQ ID NO 5427; 230pp; English.
XX CC The invention relates to plant nucleotide sequences that direct seed-,
XX CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
XX CC or constitutive transcription of an operatively linked nucleic acid
XX CC segment. The invention also relates to a method for augmenting a plant
XX CC genome and a method of identifying a gene, where its expression is
XX CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
XX CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
XX CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
XX CC sorghum, rice or wheat. The polynucleotides and the polypeptides they
XX CC encode are useful for manipulating crop plants to alter or improve
XX CC phenotypic characteristics, to produce large quantities of oil or
XX CC proteins, to incur resistance to insecticides, viruses or fungi, and to
XX CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
XX CC have a high nutritional value with reduced apical dominance or dwarfism,
XX CC early flowering or altered metabolic pathways. This sequence represents a
XX CC plant nucleic acid of the invention. Note: The sequence data for this
XX CC patent did not form part of the printed specification but was obtained in
XX CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 546 BP; 143 A; 123 C; 169 G; 111 T; 0 U; 0 Other.
XX
XX Query Match 46.9%; Score 28.6; DB 12; Length 546;
XX Best Local Similarity 72.5%; Pred. No. 16;
XX Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
XX
XX 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
XX 205 GATGAACGTTGTAAGAGCAGCAGCTCGAAGAACTATAGAAAGACTGGA 255
XX
XX RESULT 15
XX ACL33278
XX ID ACL33278 standard; cDNA, 785 BP.
XX AC ACL33278;
XX

```

```

DT 02-JUN-2005 (first entry)
XX DE Rice abiotic stress response related polynucleotide SEQ ID NO:11841.
XX KM ss; abiotic stress tolerance; transgenic plant; plant; cereal;
XX KM agriculture.
XX OS Oryza sativa.
XX PN WO2003008540-A2.
XX PD 30-JAN-2003.
XX PF 21-JUN-2002; 2002WO-US019668.
XX PR 22-JUN-2001; 2001US-0300112P.
XX PR 24-AUG-2001; 2001US-0314662P.
XX PR 26-SEP-2001; 2001US-0325277P.
XX PR 21-NOV-2001; 2001US-0332132P.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
XX PI Moughamer T, Provart N, Riche D, Zhu T;
XX DR WPI; 2003-248011/24.
XX PT New stress-responsive nucleic acid, useful for altering the
XX PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
XX PT stress, salt stress or osmotic stress.
XX PS Claim 48; SEQ ID NO 11841; 89pp; English.
XX CC The invention relates to novel abiotic stress responsive polynucleotides
XX CC and polypeptides. Also disclosed are vectors, expression cassettes, host
XX CC cells, and plants containing such polynucleotides. Also disclosed are
XX CC methods for using the polynucleotides and polypeptides to alter the
XX CC responsiveness of a plant to abiotic stress. The invention is useful in
XX CC agriculture. The nucleic acid is useful for determining whether a test
XX CC plant has been exposed to an abiotic stress condition. It is also useful
XX CC for selecting an agent that alters abiotic stress regulated
XX CC polynucleotide expression in a plant cell, and to identify a homolog or
XX CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
XX CC molecule and the polypeptide encoded by it are useful in altering the
XX CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
XX CC stress, osmotic stress or any of their combinations. The present sequence
XX CC is used in the exemplification of the invention
XX SQ Sequence 785 BP; 183 A; 198 C; 241 G; 163 T; 0 U; 0 Other;
XX
XX Query Match 46.9%; Score 28.6; DB 11; Length 785;
XX Best Local Similarity 72.5%; Pred. No. 17;
XX Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
XX
XX 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
XX 246 GATGAACGTTGTAAGAGCAGCAGCTCGAAGAACTATAGAAAGACTGGA 256
XX

```

Search completed: February 27, 2006, 07:05:24
 Job time : 40.7097 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceeleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28 ; Search time 66.7965 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtggaagaactg.....aagctgcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	61	9 US-10-789-164-4	Sequence 4, Appl1
2	46	75.4	62	9 US-10-789-164-5	Sequence 5, Appl1
3	30.4	49.8	882	6 US-10-166-653-11	Sequence 11, Appl1
4	30.4	49.8	882	10 US-11-116-192-11	Sequence 11, Appl1
5	30.2	49.5	1254	7 US-10-437-963-46191	Sequence 46191, A
6	29.6	48.5	1717	7 US-10-437-963-22178	Sequence 22178, A
7	29.4	48.2	787	7 US-09-770-445-854	Sequence 854, App
8	29.2	47.9	1372	6 US-10-369-493-28677	Sequence 28677, A
9	28.6	46.9	296	3 US-09-294-0938-324	Sequence 324, App
10	28.6	46.9	546	7 US-10-260-238-5427	Sequence 5427, App
11	28.6	46.9	785	7 US-10-260-238-4613	Sequence 4613, App
12	28.6	46.9	866	7 US-10-425-114-21330	Sequence 21330, A
13	28.6	46.9	1019	7 US-10-260-238-4614	Sequence 4614, App
14	28.6	46.9	1166	8 US-10-425-115-59277	Sequence 59277, A
15	28.6	46.9	1197	8 US-10-425-115-59276	Sequence 59276, A
16	28.6	46.9	916	6 US-10-369-493-25949	Sequence 25949, A
17	27.8	45.6	588	4 US-09-925-065A-246263	Sequence 246263, A
18	27.6	45.2	3870	10 US-11-097-143-39311	Sequence 39311, A
19	27.6	45.2	6556	10 US-11-097-143-39310	Sequence 39310, A
20	27.4	44.9	469	3 US-09-918-995-2071	Sequence 2071, App
21	27.4	44.9	1466	5 US-10-037-370-914	Sequence 914, App
22	27.4	44.9	1466	6 US-10-117-722-914	Sequence 914, App
23	27.4	44.9	1466	9 US-10-122-851-914	Sequence 914, App

24	27.4	44.9	2253	3 US-09-795-693-15	Sequence 15, Appl1
25	27.4	44.9	2253	5 US-10-156-239-15	Sequence 15, Appl1
26	27.4	44.9	2253	5 US-10-199-485-15	Sequence 15, Appl1
27	27.4	44.9	2317	6 US-10-094-749-801	Sequence 801, App
28	27.4	44.9	2593	3 US-09-795-693-13	Sequence 13, Appl1
29	27.4	44.9	2593	3 US-10-156-239-13	Sequence 13, Appl1
30	27.4	44.9	2593	5 US-10-199-485-13	Sequence 13, Appl1
31	27.4	44.9	9733	5 US-09-764-891-6690	Sequence 6890, App
32	27.4	44.9	9738	3 US-09-764-891-6691	Sequence 6891, App
33	27.4	44.9	13348	7 US-10-038-854-3	Sequence 3, Appl1
34	27.4	44.9	13935	9 US-10-450-763-10387	Sequence 10387, A
35	27.2	44.6	1057	7 US-10-425-114-24607	Sequence 24607, A
36	27.2	44.6	1146	8 US-10-425-115-59279	Sequence 59279, A
37	27	44.3	369	7 US-10-260-238-1163	Sequence 1163, App
38	27	44.3	423	7 US-10-437-963-13595	Sequence 13595, A
39	27	44.3	499	7 US-10-437-963-35811	Sequence 35811, A
40	27	44.3	588	8 US-10-425-115-4089	Sequence 4089, App
41	27	44.3	1122	7 US-10-437-963-13594	Sequence 13594, A
42	27	44.3	1518	7 US-10-437-963-135812	Sequence 6890, App
43	27	44.3	1620	6 US-10-369-493-33104	Sequence 33104, A
44	27	44.3	1707	6 US-10-369-493-32847	Sequence 32847, A
45	27	44.3	2016	6 US-10-369-493-41284	Sequence 41284, A

ALIGNMENTS

RESULT 1
US-10-789-164-4
; Sequence 4, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 4
; LENGTH: 61
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-4

Query Match 100.0%; Score 61; DB 9; Length 61;
Best Local Similarity 100.0%; Pred. No. 5e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 61; Conservative 0;

QY 1 GATCCAAGTGAAGAACTGCTCAAGAAACCGTGTCTCAAGAACTGCTCAAGAACTGT 60
DB 1 GATCCAAGTGAAGAACTGCTCAAGAAACCGTGTCTCAAGAACTGCTCAAGAACTGT 60
QY 61 A 61
DB 61 A 61

RESULT 2
US-10-789-164-5/c
; Sequence 5, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9

```
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 5
; LENGTH: 62
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-5
```

```
Query Match          75.4%; Score 46; DB 9; Length 62;
Best Local Similarity 98.3%; Pred. No. 3.3e-05;
Matches 57; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
OY      5 CAAGTGAAGAAAGCTCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB      62 CAAGTGAAGAAAGCTCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 5
```

RESULT 3

```
US-10-166-653-11
; Sequence 11, Application US/10166653
; Publication No. US20030232338A1
; GENERAL INFORMATION:
```

```
; APPLICANT: USUDA, Yoshihiro
; APPLICANT: NISHIO, Yosuke
; APPLICANT: YASUEDA, Hisashi
; APPLICANT: SUGIMOTO, Shintichi
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING POLYPEPTIDES INVOLVED IN ONE-CARBON COME
; FILE REFERENCE: 211826USO
; CURRENT APPLICATION NUMBER: US/10/166,653
; CURRENT FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Methylophilus methylotrophus
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(882)
; OTHER INFORMATION:
US-10-166-653-11
```

```
Query Match          49.8%; Score 30.4; DB 6; Length 882;
Best Local Similarity 71.4%; Pred. No. 6.5;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
OY      6 AAGTGAAGAAAGCTCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB      815 AAGAGGCTGCACCTCTGTAAGCTGCGCTCTGAGAGAGCTGCGCCAGCACTGAA 870
```

RESULT 4

```
US-11-116-192-11
; Sequence 11, Application US/11116192
; Publication No. US20050208634A1
; GENERAL INFORMATION:
```

```
; APPLICANT: USUDA, Yoshihiro
; APPLICANT: NISHIO, Yosuke
; APPLICANT: YASUEDA, Hisashi
; APPLICANT: SUGIMOTO, Shintichi
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING POLYPEPTIDES INVOLVED IN ONE-CARBON COME
; FILE REFERENCE: 211826USO
; CURRENT APPLICATION NUMBER: US/11/116,192
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/10/166,653
; PRIOR FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 882
```

```
; TYPE: DNA
; ORGANISM: Methylophilus methylotrophus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(882)
; OTHER INFORMATION:
US-11-116-192-11
```

```
Query Match          49.8%; Score 30.4; DB 10; Length 882;
Best Local Similarity 71.4%; Pred. No. 6.5;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
OY      6 AAGTGAAGAAAGCTCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB      815 AAGAGGCTGCACCTCTGTAAGCTGCGCTCTGTAAGAGCTGCGCCAGCACTGAA 870
```

RESULT 5

```
US-10-437-963-46191
; Sequence 46191, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
```

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 46191
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Oryza sativa
```

```
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MRT4530_49081C.1
US-10-437-963-46191
```

```
Query Match          49.5%; Score 30.2; DB 7; Length 1254;
Best Local Similarity 74.5%; Pred. No. 8.2;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY      11 GAAGAAAGCTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB      443 GATGAAGCTGTTGAAGAGGCGAGCTGCTGTGAAGCTATTGAAGAGCTGGA 493
```

RESULT 6

```
US-10-437-963-22178/c
; Sequence 22178, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
```

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 22178
```


APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 5427
LENGTH: 546
TYPE: DNA
ORGANISM: Zea mays
US-10-260-238-5427

Query Match 46.9%; Score 28.6; DB 7; Length 546;
Best Local Similarity 72.5%; Pred. No. 22;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 205 GATGAACGTGTGAAGAGCAGCAGCTCGAAGAACTATAGAAGAGCTGGA 255

RESULT 11

Sequence 4613, Application US/10260238
Publication No. US20040016025A1
GENERAL INFORMATION:
APPLICANT: Budworth, Paul R.
APPLICANT: Moughamer, Todd G.
APPLICANT: Briggs, Steven P.
APPLICANT: Cooper, Bret
APPLICANT: Glazebrook, Jane
APPLICANT: Goff, Stephen A.
APPLICANT: Katagiri, Fumiyaki
APPLICANT: Kreps, Joel
APPLICANT: Provart, Nicholas
APPLICANT: Rieke, Darrell
APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 4613
LENGTH: 785
TYPE: DNA
ORGANISM: Trifolium aestivum
US-10-260-238-4613

Query Match 46.9%; Score 28.6; DB 7; Length 785;
Best Local Similarity 72.5%; Pred. No. 24;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 246 GATGAACGTGTGAAGAGCAGCAGCTCGAAGAACTATAGAAGAGCTGGA 296

RESULT 12

US-10-425-114-21330
Sequence 21330, Application US/10425114
Publication No. US20040034888A1

GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yinhua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tadaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 21330
LENGTH: 866
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3279-232-C9_FLI
US-10-425-114-21330

Query Match 46.9%; Score 28.6; DB 7; Length 866;
Best Local Similarity 72.5%; Pred. No. 25;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 346 GATGAACGTGTGAAGAGCAGCAGCTCGAAGAACTATAGAAGAGCTGGA 396

RESULT 13

Sequence 4614, Application US/10260238
Publication No. US20040016025A1
GENERAL INFORMATION:
APPLICANT: Budworth, Paul R.
APPLICANT: Moughamer, Todd G.
APPLICANT: Briggs, Steven P.
APPLICANT: Cooper, Bret
APPLICANT: Glazebrook, Jane
APPLICANT: Goff, Stephen A.
APPLICANT: Katagiri, Fumiyaki
APPLICANT: Kreps, Joel
APPLICANT: Provart, Nicholas
APPLICANT: Rieke, Darrell
APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 4614
LENGTH: 1019
TYPE: DNA
ORGANISM: Trifolium aestivum
FEATURE:
NAME/KEY: N region
LOCATION: (31)..(31)
OTHER INFORMATION: n = any nucleotide
FEATURE:
NAME/KEY: N region
LOCATION: (117)..(136)
OTHER INFORMATION: n = any nucleotide
US-10-260-238-4614

Query Match 46.9%; Score 28.6; DB 7; Length 1019;
Best Local Similarity 72.5%; Pred. No. 26;

Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 349 GATGAAACGTTGAAACAAGCAGCTGCTCAAGAACTATAGAAAGAGCTGA 399

RESULT 14

US-10-425-115-59277
; Sequence 59277, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 59277
; LENGTH: 1166
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154057C.1
US-10-425-115-59277

Query Match 46.9%; Score 28.6; DB 8; Length 1166;
Best Local Similarity 72.5%; Pred. No. 26;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 428 GATGAAACGTTGAAAGAGCAGCTCGAAGACTATAGAAAGAGCTGGA 478

RESULT 15

US-10-425-115-59276
; Sequence 59276, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 59276
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1197)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154056C.1
US-10-425-115-59276

Query Match 46.9%; Score 28.6; DB 8; Length 1197;
Best Local Similarity 72.5%; Pred. No. 27;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 474 GATGAAACGTTGAAAGAGCAGCTCGAAGACTATAGAAAGAGCTGGA 524

Search completed: February 27, 2006, 07:58:44
Job time: 67.7965 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 36.7602 Seconds
(without alignments) 3538.457 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtgcgaagaactg.....aagctgcctcaagaactgtga 61

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA New:

1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	27.8	45.6	588	US-09-925-065A-246263	Sequence 246263, Appl1
2	27.4	44.9	2654	US-10-505-263-1	Sequence 1, Appl1
3	27.4	44.9	2748	US-10-505-263-3	Sequence 3, Appl1
4	26.4	43.3	360	US-09-925-065A-566617	Sequence 566617, A
5	26.4	43.3	360	US-09-925-065A-566619	Sequence 566619, A
6	26.4	43.3	360	US-09-925-065A-566618	Sequence 566618, A
7	25.8	42.3	575	US-09-925-065A-318258	Sequence 318258, A
8	25.6	42.0	793	US-10-750-185-55250	Sequence 55250, A
9	25.6	42.0	793	US-10-750-623-55250	Sequence 55250, A
10	25.4	41.6	3388	US-11-077-550-41	Sequence 41, Appl1
11	25.4	41.6	3388	US-11-077-550-47	Sequence 47, Appl1
12	25.4	41.6	3388	US-11-077-550-51	Sequence 51, Appl1
13	25.4	41.6	3388	US-11-077-550-55	Sequence 55, Appl1
14	25.4	41.6	3397	US-11-077-550-45	Sequence 45, Appl1
15	25.4	41.6	5292	US-10-775-169-330	Sequence 330, App
16	25.2	41.3	1175	US-11-102-026A-111	Sequence 111, App
17	25.2	41.3	1215	US-11-102-026A-187	Sequence 187, App
18	25.2	41.3	108214	US-11-117-187-211	Sequence 211, App
19	25	41.0	1008	US-11-196-475-130	Sequence 130, App
20	25	41.0	1008	US-11-196-475-131	Sequence 131, App

21	25	41.0	1011	US-11-196-475-51	Sequence 51, Appl1
22	25	41.0	2460	US-11-072-512-37	Sequence 37, Appl1
23	24.8	40.7	522	US-09-925-065A-367960	Sequence 367960, A
24	24.8	40.7	1226	US-10-750-185-54946	Sequence 54946, A
25	24.8	40.7	1226	US-10-750-623-54946	Sequence 54946, A
26	24.8	40.7	1240	US-09-925-065A-553111	Sequence 553111, A
27	24.6	40.3	390	US-09-925-065A-154591	Sequence 154591, A
28	24.6	40.3	2430	US-10-649-457-3	Sequence 3, Appl1
29	24.4	40.0	200	US-11-098-666-6643	Sequence 6643, Ap
30	24.4	40.0	603	US-09-925-065A-500955	Sequence 500955, A
31	24.4	40.0	603	US-09-925-065A-500956	Sequence 500956, A
32	24.4	40.0	677	US-09-925-065A-92084	Sequence 92084, A
33	24.4	40.0	1251	US-11-098-686-9738	Sequence 9738, Ap
34	24.4	40.0	1457619	US-11-098-686-8739	Sequence 8739, Ap
35	24.2	39.7	609	US-10-750-185-46785	Sequence 46785, A
36	24.2	39.7	609	US-10-750-623-46785	Sequence 46785, A
37	24.2	39.7	643	US-09-925-065A-801762	Sequence 801762, A
38	24.2	39.7	643	US-09-925-065A-801763	Sequence 801763, A
39	24	39.3	1008	US-11-196-475-53	Sequence 53, Appl1
40	24	39.3	1773	US-11-040-218-12	Sequence 12, Appl1
41	24	39.3	2557	US-11-121-086-49	Sequence 49, Appl1
42	24	39.3	159146	US-10-775-169-77	Sequence 77, Appl1
43	23.8	39.0	507	US-09-925-065A-228106	Sequence 228106, A
44	23.8	39.0	577	US-09-925-065A-888138	Sequence 888138, A
45	23.8	39.0	589	US-09-925-065A-160795	Sequence 160795, A

ALIGNMENTS

RESULT 1
US-09-925-065A-246263
; Sequence 246263, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Manq, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 246263
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-925-065A-246263
Query Match 45.6%; Score 27.8; DB 6; Length 588;
Best Local Similarity 74.5%; Pred. No. 2.5; Indels 0; Gaps 0;
Matches 35; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 6 AACTGAGAGAACTGCTCAAGAAACCGTCTCAAGAACTGCTCA 52
Db 305 ATGTGTAAGAGGTGCCCAAGAACTGCTGTAATTAAGCTATGAA 351
RESULT 2
US-10-505-263-1
; Sequence 1, Application US/10505263
; Publication No. US20060014940A1
; GENERAL INFORMATION:

```

; APPLICANT: Vanderbilt University
; APPLICANT: Case Western Reserve University
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mount, David B
; APPLICANT: Romero, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF SLC26A6, SLC26A1, and SLC26A2
; FILE REFERENCE: 1242/50/2 PCT/US
; CURRENT FILING DATE: 2004-08-20
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2654
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-263-1

Query Match          44.9%; Score 27.4; DB 7; Length 2654;
Best Local Similarity 75.6%; Pred. No. 5;
Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
Qy      3  TCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTG 47
Db      1778 TCCCAAGAAAGAACTGCTCAAGAAAGCAGACGAGCTGAAAGCTG 1822
```

```

RESULT 3
US-10-505-263-3
; Sequence 3, Application US/10505263
; Publication No. US20060014940A1
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Case Western Reserve University
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mount, David B
; APPLICANT: Romero, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF SLC26A6, SLC26A1, and SLC26A2
; FILE REFERENCE: 1242/50/2 PCT/US
; CURRENT FILING DATE: 2004-08-20
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 2748
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (252)..(2468)
US-10-505-263-3
```

```

Query Match          44.9%; Score 27.4; DB 7; Length 2748;
Best Local Similarity 75.6%; Pred. No. 5.1;
Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
Qy      3  TCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTG 47
Db      1908 TCCCAAGAAAGAACTGCTCAAGAAAGCAGACGAGCTGAAAGCTG 1952
```

```

RESULT 4
US-09-925-065A-566617
; Sequence 566617, Application US/09925065A
```

```

; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; CURRENT FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 566617
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-566617

Query Match          43.3%; Score 26.4; DB 6; Length 360;
Best Local Similarity 69.2%; Pred. No. 6.9;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
Qy      10  GGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
Db      140  GGAGGAGCTGTTACAAAACAGCTTCCAGGCACTGCTCGAATCTGTA 191
```

```

RESULT 5
US-09-925-065A-566618
; Sequence 566618, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; CURRENT FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 566618
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-566618
```

```

Query Match          43.3%; Score 26.4; DB 6; Length 360;
Best Local Similarity 69.2%; Pred. No. 6.9;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
Qy      10  GGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
Db      140  GGAGGAGCTGTTACAAAACAGCTTCCAGGCACTGCTCGAATCTGTA 191
```

RESULT 6

US-09-925-065A-566619

; Sequence 566619, Application US/09925065A

; Publication No. US20040181048A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single

; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925,065A

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 566619

; LENGTH: 360

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-566619

Query Match 43.3%; Score 26.4; DB 6; Length 360;

Best Local Similarity 69.2%; Pred. No. 6.9; Mismatches 16; Indels 0; Gaps 0;

Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

10 GGAGAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61

140 GGAGAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTGGAATCTGTA 191

RESULT 7

US-09-925-065A-318258

; Sequence 318258, Application US/09925065A

; Publication No. US20040181048A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single

; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925,065A

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 318258

; LENGTH: 575

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-318258

Query Match 42.3%; Score 25.8; DB 6; Length 575;

Best Local Similarity 67.9%; Pred. No. 13; Mismatches 17; Indels 0; Gaps 0;

Matches 36; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

6 AAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58

Db 232 AAGTGTGTGTAACACCTGAGAAAAACGTTGTGAAAAAGACGCTGAAGAACT 284

RESULT 8

US-10-750-185-55250/c

; Sequence 55250, Application US/10750185

; Publication No. US20050260603A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KERR, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-2

; CURRENT APPLICATION NUMBER: US/10/750,185

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO: 55250

; LENGTH: 793

; TYPE: DNA

; ORGANISM: Bovine

US-10-750-185-55250

Query Match 42.0%; Score 25.6; DB 8; Length 793;

Best Local Similarity 70.8%; Pred. No. 16; Mismatches 14; Indels 0; Gaps 0;

Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

2 ATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCT 49

142 ATCCAAGTGAAGAACTGCTGAGAACTGCTGCAAGCTGCTGCAAGCTGCT 95

RESULT 9

US-10-750-623-55250/c

; Sequence 55250, Application US/10750623

; Publication No. US20050287531A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KERR, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis

; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-1

; CURRENT APPLICATION NUMBER: US/10/750,623

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO: 55250

; LENGTH: 793

; TYPE: DNA

; ORGANISM: Bovine

US-10-750-623-55250

Query Match 42.0%; Score 25.6; DB 8; Length 793;

Best Local Similarity 70.8%; Pred. No. 16; Mismatches 14; Indels 0; Gaps 0;

Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

2 ATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCT 49

142 ATCCAAGTGAAGAACTGCTGAGAACTGCTGCAAGCTGCTGCAAGCTGCT 95


```

RESULT 10
US-11-077-550-41
; Sequence 41, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foester, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077.550
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GS97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-11-077-550-41

Query Match      41.6%  Score 25.4;  DB 11;  Length 3388;
Beet Local Similarity 64.4%  Pred. No. 27;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0.

Cy      2  ATCCGAGGGAAGAAGACTGCTCAAGAAACCGCTGCTCAAGAAAGCGCTCAAGAACTG 60
Db      2608 ATCAAGGCTGAAGCGCGCGCTTAAGAAAGCGCGCTTAAGAAAGCG3CGCTTAAGCGGT 2666

RESULT 11
US-11-077-550-47
; Sequence 47, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foester, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077.550
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GS97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-11-077-550-41

```

```

; PRIOR APPLICATION NUMBER: GB9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin version 3.1
; SEQ ID NO: 47
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-11-077-550-47

Query Match      41.6%; Score 25.4; DB 11; Length 3388;
Best Local Similarity    64.4%; Pred. No. 27;
Matches   38; Conservative    0; Mismatches   21; Indels    0; Gaps    0;

Oy       2 ATCCAAGTGGAGAAGAACTGCTCAAGAACCGCCTGCTCAAGAAAGTGTGCTCAAGAAACTGT 60
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       2608 ATCAAAGCTGAAGCGGCCGCTTAAGAAAGCGCCGCTAAGAAAGCGCCGCTAAGACGCT 2666


RESULT 12
US-11-077-550-51
; Sequence 51, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chadock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077,550
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin version 3.1
; SEQ ID NO: 51
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-11-077-550-51

Query Match      41.6%; Score 25.4; DB 11; Length 3388;
Best Local Similarity    64.4%; Pred. No. 27;
Matches   38; Conservative    0; Mismatches   21; Indels    0; Gaps    0;

Oy       2 ATCCAAGTGGAGAAGAACTGCTCAAGAACCGCCTGCTCAAGAAAGTGTGCTCAAGAAACTGT 60
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       2608 ATCAAAGCTGAAGCGGCCGCTTAAGAAAGCGCCGCTAAGAAAGCGCCGCTAAGACGCT 2666


RESULT 13
US-11-077-550-55
; Sequence 55, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 14.3542 Seconds
(without alignments)
7553.984 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtcgagaagaactg.....aagctgcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/1 COMB.seq:*
2: /cgn2_6/prodata/1/ina/5 COMB.seq:*
3: /cgn2_6/prodata/1/ina/6a COMB.seq:*
4: /cgn2_6/prodata/1/ina/6b COMB.seq:*
5: /cgn2_6/prodata/1/ina/H COMB.seq:*
6: /cgn2_6/prodata/1/ina/RTCTUS COMB.seq:*
7: /cgn2_6/prodata/1/ina/PP COMB.seq:*
8: /cgn2_6/prodata/1/ina/RE COMB.seq:*
9: /cgn2_6/prodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	30.4	49.8	882 3 US-10-166-653-11	Sequence 11, Appl
2	28.4	46.6	354 3 US-09-248-796A-5595	Sequence 5595, Ap
3	28.2	46.2	1920 2 US-08-186-222-1	Sequence 1, Appl1
4	27.8	45.6	2103 3 US-09-107-532A-1440	Sequence 1440, Ap
5	27.4	44.9	1466 3 US-09-620-312D-914	Sequence 914, App
6	27.2	44.6	47115 3 US-09-949-016-12278	Sequence 12278, A
7	27.2	44.6	47122 3 US-09-949-016-16520	Sequence 16520, A
8	26.8	43.9	5962 9 5386025-5	Patent No. 5386025
9	26.8	43.9	5975 2 US-08-404-354B-1	Sequence 1, Appl1
10	26.8	43.9	5975 2 US-08-314-083B-1	Sequence 1, Appl1
11	26.8	43.9	5975 2 US-08-435-675B-1	Sequence 1, Appl1
12	26.8	43.9	5975 2 US-08-336-257A-3	Sequence 3, Appl1
13	26.8	43.9	5975 2 US-08-884-599-1	Sequence 1, Appl1
14	26.6	43.6	1140 3 US-09-328-352-1093	Sequence 1093, Ap
15	26.4	43.3	324 3 US-09-902-540-7007	Sequence 7007, Ap
16	26.4	43.3	5164 3 US-09-902-540-615	Sequence 615, App
17	26.2	43.0	516 3 US-09-270-767-753B	Sequence 753B, App
18	26.2	43.0	516 3 US-09-270-767-22820	Sequence 22820, A
19	26.2	43.0	723 3 US-09-603-208A-9	Sequence 9, Appl1
20	26.2	43.0	1548 3 US-09-248-796A-6167	Sequence 6167, Ap
21	25.6	42.0	528 3 US-09-248-796A-4819	Sequence 4819, Ap
22	25.6	42.0	1131 3 US-09-248-796A-994	Sequence 994, App
23	25.2	41.3	345 3 US-09-134-000C-1925	Sequence 1925, Ap
24	25.2	41.3	1658 3 US-09-016-434-1320	Sequence 1320, Ap

25	25.2	41.3	2843 3 US-09-620-312D-211	Sequence 211, App
26	25.2	41.3	3218 3 US-09-799-451-470	Sequence 470, App
27	25.2	41.3	7176 3 US-09-221-017B-626	Sequence 626, App
28	25.2	41.3	1230025 3 US-09-198-452A-1	Sequence 1, Appl1
29	25.2	41.3	1230230 3 US-09-438-185A-1	Sequence 1092, Ap
30	25	41.0	912 3 US-09-540-236-1092	Sequence 130, App
31	25	41.0	1008 3 US-08-235-836C-130	Sequence 130, App
32	25	41.0	1008 3 US-08-235-836C-131	Sequence 131, App
33	25	41.0	1011 3 US-08-235-836C-51	Sequence 51, Appl
34	25	41.0	1548 3 US-09-799-451-625	Sequence 625, App
35	25	41.0	1945 2 US-08-525-742-5	Sequence 5, Appl1
36	25	41.0	2430 3 US-09-252-991A-12709	Sequence 12709, A
37	25	41.0	2460 3 US-10-104-047-37	Sequence 37, App
38	25	41.0	2607 3 US-09-252-991A-13249	Sequence 13249, A
39	24.8	40.7	501 3 US-09-270-767-15621	Sequence 15621, A
40	24.8	40.7	501 3 US-09-270-767-15621	Sequence 19228, A
41	24.6	40.3	200 3 US-09-513-999C-19228	Sequence 3082, Ap
42	24.6	40.3	375 3 US-09-134-000C-3082	Sequence 3, Appl1
43	24.6	40.3	1209 3 US-09-464-483-3	Sequence 3, Appl1
44	24.6	40.3	1209 3 US-09-414-664-3	Sequence 1, Appl1
45	24.6	40.3	1589 3 US-09-464-483-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-10-166-653-11
Sequence 11, Application US/1016653
Patent No. 6911332
GENERAL INFORMATION:
APPLICANT: USUDA, Yoshihiro
APPLICANT: NISHIO, Yosuke
APPLICANT: YASUEDA, Hisashi
APPLICANT: SUGIMOTO, Shinichi
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING POLYPEPTIDES INVOLVED IN ONE-CARBON COMPOUND METABOLISM IN METHYLOPHILUS METHYLOTROPHUS
FILE REFERENCE: 211926USO
CURRENT APPLICATION NUMBER: US/10166, 653
CURRENT FILING DATE: 2002-06-12
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Patentin version 3.1
SEQ ID NO 11
LENGTH: 882
TYPE: DNA
ORGANISM: Methylophilus methylotrophus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(882)
OTHER INFORMATION:
US-10-166-653-11
Query Match 49.8%; Score 30.4; DB 3; Length 882;
Best Local Similarity 71.4%; Pred. No. 0.89; Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Oy 6 AACTGGAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 815 AAGAGCTGACCTGCTGAAGCTGCGCTGCTGAAGAACTGCGCCAGCAACTGAA 870
RESULT 2
US-09-248-796A-5595
Sequence 5595, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248, 796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074, 725

PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 5595
LENGTH: 354
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-5595

Query Match 46.6%; Score 28.4; DB 3; Length 354;
Best Local Similarity 76.1%; Pred. No. 3.3;
Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 56
DB 286 GAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 331

RESULT 3

US-08-186-222-1
Sequence 1, Application US/08186222
Patent No. 5559007

GENERAL INFORMATION:

APPLICANT: Surti, Bruno

APPLICANT: Schmitz, Albert

TITLE OF INVENTION: Bacterial Vectors

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: New York

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/186,222

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/672,205

FILING DATE: 19-MAR-1991

APPLICATION NUMBER: GB 9006400.7

FILING DATE: 22-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Villamizar, Joann

REGISTRATION NUMBER: 30,598

REFERENCE/DOCKET NUMBER: 4-17994/A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (914)785-7121

TELEFAX: (914)347-5769

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1920 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Lactococcus lactis LM0230

INDIVIDUAL ISOLATE: Major Secretion Product (MSP) Gene

IMMEDIATE SOURCE:

CLONE: pUCS

FEATURE:

NAME/KEY: CDS

LOCATION: 411..1793

US-08-186-222-1

Query Match 46.2%; Score 28.2; DB 2; Length 1920;
Best Local Similarity 73.5%; Pred. No. 5.8;
Matches 36; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 7 AGTGAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 55
DB 1061 AGATGAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 1109

RESULT 4

US-09-107-532A-1440
Sequence 1440, Application US/09107532A
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-8277

TELEFAX: (781)893-5007

INFORMATION FOR SEQ ID NO: 1440:

SEQUENCE CHARACTERISTICS:

LENGTH: 2103 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc_feature

LOCATION: (B) LOCATION 1..2103

SEQUENCE DESCRIPTION: SEQ ID NO: 1440:

US-09-107-532A-1440

Query Match 45.6%; Score 27.8; DB 3; Length 2103;
Best Local Similarity 69.1%; Pred. No. 8;
Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 5 CAAGTGAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAACTG 59
DB 1687 CAAGTGAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAACTG 1741

RESULT 5

US-09-620-312D-914

Sequence 914, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenchua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhimei
APPLICANT: John, Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE REFERENCE: 784C1P2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_fl_genes Version 1.0
SEQ ID NO 914
LENGTH: 1466
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (779)..(1213)
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(1466)
OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-914

Query Match 44.9%; Score 27.4; DB 3; Length 1466;
Best Local Similarity 75.6%; Pred. No. 9; 9;
Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 3 TCCTAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTG 47
Db 418 TCCTAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTG 462

RESULT 6
US-09-949-016-12278/c
Sequence 12278, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12278

LENGTH: 47115
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(47115)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12278

Query Match 44.6%; Score 27.2; DB 3; Length 47115;
Best Local Similarity 67.9%; Pred. No. 28;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 56
Db 39716 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTGCTCAACAGATGCTCAACAA 39661

RESULT 7
US-09-949-016-16520/c
Sequence 16520, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16520
LENGTH: 47122
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(47122)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16520

Query Match 44.6%; Score 27.2; DB 3; Length 47122;
Best Local Similarity 67.9%; Pred. No. 28;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 56
Db 39716 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTGCTCAACAGATGCTCAACAA 39661

RESULT 8
US-09-949-016-16520/c
Sequence 16520, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16520
LENGTH: 47122
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(47122)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16520

Query Match 43.9%; Score 26.8; DB 9; Length 5962;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GTGGAAGAACTGCTCAAGAAACGCTGCTCAAGAGCTGCTCAAG 53
DB 111 GAGGAAGAAACAGCCCAAGAGCCCTGCTCCGAGGCTCTGCCCCAGG 156

RESULT 9

US-08-404-354B-1

Sequence 1, Application US/08404354B.

Patent No. 5618720

GENERAL INFORMATION:

APPLICANT: Ellis, Steven Bradley

APPLICANT: Williams, Mark E.

APPLICANT: Harpold, Michael Miller

APPLICANT: Schwartz, Robert

TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brown, Martin, Haller & McClain

STREET: 1660 Union Street

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92101-2926

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/404,354B

FILING DATE: 15-FEB-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,083

FILING DATE: 28-SEP-1994

APPLICATION NUMBER: US 07/914,231

FILING DATE: 13-JUL-1992

APPLICATION NUMBER: US 07/603,751

FILING DATE: 08-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-53192

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5975 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 79...5700

OTHER INFORMATION:

US-08-404-354B-1

Query Match 43.9%; Score 26.8; DB 2; Length 5975;

Best Local Similarity 73.9%; Pred. No. 22;

Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GTGGAAGAACTGCTCAAGAAACGCTGCTCAAGAGCTGCTCAAG 53

DB 111 GAGGAAGAAACAGCCCAAGAGCCCTGCTCCGAGGCTCTGCCCCAGG 156

RESULT 10

US-08-314-083B-1

Sequence 1, Application US/08314083B
Patent No. 5686241

GENERAL INFORMATION:

APPLICANT: Ellis, Steven Bradley

APPLICANT: Williams, Mark E.

APPLICANT: Harpold, Michael Miller

APPLICANT: Schwartz, Robert

TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brown, Martin, Haller & McClain

STREET: 1660 Union Street

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92101-2926

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/314,083B

FILING DATE: 28-SEP-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/914,231

FILING DATE: 13-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/603,751

FILING DATE: 08-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-53191

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5975 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 79...5700

OTHER INFORMATION:

US-08-314-083B-1

Query Match 43.9%; Score 26.8; DB 2; Length 5975;

Best Local Similarity 73.9%; Pred. No. 22;

Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GTGGAAGAACTGCTCAAGAAACGCTGCTCAAGAGCTGCTCAAG 53

DB 111 GAGGAAGAAACAGCCCAAGAGCCCTGCTCCGAGGCTCTGCCCCAGG 156

RESULT 11

US-08-435-675B-1

Sequence 1, Application US/08435675B

Patent No. 5710250

GENERAL INFORMATION:

APPLICANT: Ellis, Steven Bradley

APPLICANT: Williams, Mark E.

APPLICANT: Harpold, Michael Miller

APPLICANT: Schwartz, Robert

TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435, 675B
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,083
FILING DATE: 28-SEP-1994
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53193
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79..5700
OTHER INFORMATION:
US-08-435-675B-1
Query Match 43.9%; Score 26.8; DB 2; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 8 GTGGAAGAACTGCTCAAGAAACCGCTGCTCAAGAAAGCTGCTCAAG 53
DB 111 GAGGAAGAAACAGCCCAAGAAAGCCCTGCGCCGAGTCTCCAGG 156
RESULT 12
US-08-336-257A-3
Sequence 3, Application US/08336257A
Patent No. 5726035
GENERAL INFORMATION:
APPLICANT: Jay, Scott D
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
APPLICANT: Campbell, Kevin P.
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336, 257A
FILING DATE: 07-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 54898
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79..5700
OTHER INFORMATION: "product= "Alpha-1 subunit of animal calcium
channel"
US-08-336-257A-3
Query Match 43.9%; Score 26.8; DB 2; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 8 GTGGAAGAACTGCTCAAGAAACCGCTGCTCAAGAAAGCTGCTCAAG 53
DB 111 GAGGAAGAAACAGCCCAAGAAAGCCCTGCGCCGAGTCTCCAGG 156
RESULT 13
US-08-884-599-1
Sequence 1, Application US/08884599
Patent No. 6013474
GENERAL INFORMATION:
APPLICANT: Williams, Steven Bradley
APPLICANT: Williams, Mark B.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Arnold
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/884, 599
FILING DATE: 27-JUNE-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/314,083
FILING DATE: 28-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751

FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53191B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79..5700
OTHER INFORMATION:
US-08-884-599-1

Query Match 43.9%; Score 26.8; DB 3; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 8 GTGGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAG 53
DB 111 GAGGAAGAAACGACCCCAAGAGCCCTGCGAGGCTGCTGCGCCAG 156

RESULT 14
US-09-328-352-1093
Sequence 1093, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 1093
LENGTH: 1140
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-1093

Query Match 43.6%; Score 26.6; DB 3; Length 1140;
Best Local Similarity 71.4%; Pred. No. 17;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 7 AGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 55
DB 1092 AGTACTCAACTGCTCAAGAGCGGCACTCTCTGACGACTCAATTA 1140

RESULT 15
US-09-902-540-7007
Sequence 7007, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(13849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10

NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 7007
LENGTH: 324
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-7007

Query Match 43.3%; Score 26.4; DB 3; Length 324;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 12 AAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 55
DB 52 AAGCAAGAGTCAAGAAACCGCGCTCTACAGAGTGTCTGCA 95

Search completed: February 27, 2006, 06:46:21
Job time : 16.5542 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 368.475 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaagcgtctc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sta:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB. ID	Description
C 1	28.8	45.0	86313	5	CR558305 Zebrafish
C 2	28.8	45.0	213324	9	AC161816 Mus muscu
C 3	28.6	44.7	25600	2	U80442 Caenorhabdi
C 4	28.6	44.7	189293	9	AC161379 Mus muscu
C 5	28.6	44.7	201463	14	AC154577 Mus muscu
C 6	28.6	44.7	280474	14	AC160041 Bos tauru
C 7	28.6	44.7	294136	14	AC006901 Caenorhab
C 8	28.6	44.4	172866	14	AC164240 Bos tauru
C 9	28.2	44.1	145870	14	AC149957 Strongylo
C 10	28.2	44.1	229461	14	AC155077 Bos tauru
C 11	27.8	43.4	196622	8	AC125238 Homo sapi
C 12	27.4	42.8	81826	8	HSU78045 Homo colia
C 13	27.4	42.8	173359	14	AP000922 Homo sapi
C 14	27.4	42.8	180399	14	AP001144 Homo sapi
C 15	27.2	42.5	34337	6	AX398707 Sequence
C 16	27.2	42.5	61550	8	AY049008 Homo sapi
C 17	27.2	42.5	87187	8	AL161744 Human DNA
C 18	27.2	42.5	110000	14	AC055726_2 Continuation (3 of

19	27.2	42.5	148390	9	AC121930 Mus muscu
C 20	27.2	42.5	160008	14	AC022823 Human sapi
C 21	27.2	42.5	162228	8	AL450382 Human DNA
C 22	27.2	42.5	163231	8	AL954360 Human DNA
C 23	27.2	42.5	166256	8	AC018922 Homo sapi
C 24	27.2	42.5	167071	8	AC128649 Homo sapi
C 25	27.2	42.5	167130	14	AL583858 Homo sapi
C 26	27.2	42.5	255976	9	AC147230 Mus muscu
C 27	27.2	42.5	274996	14	AC096298 Rattus no
C 28	27.2	42.5	324582	14	AC133362 Rattus no
C 29	26.8	41.9	10841	1	AE013054 Thermococ
C 30	26.8	41.9	59818	4	AC123967 Smittoph
C 31	26.6	41.6	1945	5	AB196462 Oncorhyn
C 32	26.6	41.6	110000	15	AE016818_01 Continuation (2 of
C 33	26.6	41.6	169794	14	AC004688 Plasmodiu
C 34	26.6	41.6	172883	9	AC132274 Mus muscu
C 35	26.6	41.6	196149	14	AC004709 Plasmodiu
C 36	26.6	41.6	252650	2	AE014847 Plasmodiu
C 37	26.4	41.2	1209	15	AK108702 Oryza sat
C 38	26.4	41.2	1578	6	AX834701 Sequence
C 39	26.4	41.2	1578	8	AK097338 Homo sapi
C 40	26.4	41.2	2434	8	AF125442 Caenorhab
C 41	26.4	41.2	31818	2	AF125442 Caenorhab
C 42	26.4	41.2	110000	15	AP008211_237 Continuation (23 of
C 43	26.4	41.2	127642	15	AC093952 Oryza sat
C 44	26.4	41.2	137357	15	AC118289 Oryza sat
C 45	26.2	40.9	5673	15	AY026257 Magnapor

ALIGNMENTS

RESULT 1
CR558305/c
LOCUS
DEFINITION
Zebrafish DNA sequence from clone DKRP-87C9 in linkage group 13,
complete sequence.
ACCESSION
CR558305.6 GI:55294877
VERSION
HTG.
KEYWORDS
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (03-NOV-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requesters: clonerequest@sanger.ac.uk
On Nov 3, 2004 this sequence version replaced gi:54606595.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:
 Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
 Zebrafish pUC subclones occasionally display inconsistency over the
 length of mononucleotide A/T runs and conserved TA repeats. Where
 this is found the longest good quality representation will be
 submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat
 discovery system (Zhifeng Bao and Sean Eddy, submitted), and those
 beginning 'dir' were identified by Rick Waterman (Stephen Johnson
 lab, Washu). For further information see
http://www.sanger.ac.uk/Projects/D_reio/fishmask.shtml DKEYP-87C9
 is from a Zebrafish BAC library
 VECTOR: pindigobAC-5

FEATURES
 Source
 Location/Qualifiers
 1. 86313

/organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEYP-87C9"
 /clone_lib="DanioKeypl10c"

ORIGIN

Query Match 45.0%; Score 28.8; DB 5; Length 86313;
 Best Local Similarity 82.5%; Pred. No. 57;
 Matches 33; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

22 AACGCTGAGAGAGCTTTAGCAAAATCTGGAAGTGGAA 61
 68274 AACGCTGAGAGATTTGTTTCAAACTCTGATCTTGA 68235

RESULT 2
 AC161816 213324 bp DNA linear ROD 03-AUG-2005
 LOCUS AC161816
 DEFINITION Mus musculus BAC clone RP23-308L3 from chromosome 16, complete
 sequence.

AC161816
 AC161816.5 GI:71725592
 HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Tomlinson, C., Lewis, S. and Levy, A.
 TITLE The sequence of Mus musculus BAC clone RP23-308L3.
 JOURNAL Unpublished (2001)
 REFERENCE
 AUTHORS Wilson, R.K.
 TITLE Direct Submision
 JOURNAL Submitted (20-MAY-2005) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Wilson, R.K.
 TITLE Direct Submision
 JOURNAL Submitted (23-JUN-2005) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Wilson, R.K.
 TITLE Direct Submision
 JOURNAL Submitted (03-AUG-2005) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT On Aug 3, 2005 this sequence version replaced gi:68159160.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 ----- Summary Statistics

Center project name: M_BA0308L03

NOTICE:

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e. phred quality
 >=30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone, fosmid clone or direct clone walk sequence.
 Sequence from the Mouse Genome Sequencing Consortium whole genome
 shotgun may have been used to obtain a consensus sequence. The
 assembly was confirmed by restriction digest.
 This finishing standard has slightly changed from the previous
 Human standard. Specifically, standards for regions of low sequence
 complexity (such as dinucleotide repeats and small unit tandem
 repeats) have been relaxed. These regions are very prevalent in the
 mouse genome, and the return on extended finishing efforts is
 minimal.

If a sequence meets the criteria of the above statement, it needs
 no comments or tags. If the criteria are not met, such as ambiguous
 bases, then the region is duly annotated.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. Wes Warren,
 Department of Genetics, Washington University, St. Louis MO. For
 additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:
 The BAC library has been constructed by Kazuhiro Oseegawa and
 Minako Tateo in the laboratory of Pieter de Jong
 (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or
 brain genomic DNA. The clone and detailed information can be
 obtained from Research Genetics, Inc. (<http://www.resgen.com>) or
 Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

FEATURES

Source

Location/Qualifiers
 1. 213324
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="16"
 /clone="RP23-308L3"
 /clone_lib="RPCT-23"

misc_feature
 144936..145884
 /note="Sequence derived from PCR product of project DNA"
 unsure
 147247..147321
 /note="Sequence derived from one plasmid subclone."
 unsure
 148786..148957
 /note="Sequence derived from one plasmid subclone."
 unsure
 150162..150279
 /note="Sequence derived from one plasmid subclone."
 unsure
 151242..151320
 /note="Sequence derived from one plasmid subclone."
 unsure
 161365..161388
 /note="Sequence derived from one plasmid subclone."
 unsure
 172382..172405
 /note="Sequence derived from one plasmid subclone."
 unsure
 174673..174686
 /note="Sequence derived from one plasmid subclone."
 misc_feature
 175130..175287
 /note="Sequence derived from PCR product of project DNA"
 misc_feature
 206133..206193
 /note="Sequence derived from one plasmid subclone."
 unsure

ORIGIN

Query Match 45.0%; Score 28.8; DB 9; Length 213324;
 Best Local Similarity 65.6%; Pred. No. 59;
 Matches 42; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 1 GATCCGCAAGAGTGTTCACACCGCTGAGAGCTGTAGCAATCTGCACTGGA 60
 DB 190239 GGTCCACAAAGAGGTGTGAGATTCACGAGAGAGTATTCTAAAGATCAAGGTCTGGA 190298
 OY 61 AGTA 64
 DB 190239 TCTA 190302
 RESULT 3
 U80442/c
 LOCUS U80442 25600 bp DNA linear INV 22-SEP-2004
 DEFINITION Caenorhabditis elegans cosmid T20F5, complete sequence.
 ACCESSION U80442
 VERSION U80442.1 GI:1703614
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 25600)
 REFERENCE
 AUTHORS WormBase Consortium
 CONSRTM Genome sequence of the nematode C. elegans: a platform for
 TITLE investigating biology. The C. elegans Sequencing Consortium
 JOURNAL Science 282 (5396), 2012-2018 (1998)
 PUBMED 9851916
 REFERENCE
 AUTHORS 2 (bases 1 to 25600)
 TITLE Du, Z. and Le T.T.
 JOURNAL The sequence of C. elegans cosmid T20F5
 TITLE Unpublished (2001)
 REFERENCE
 AUTHORS 3 (bases 1 to 25600)
 TITLE Waterston, R.
 JOURNAL Direct Submission
 TITLE Submitted (27-NOV-1996) Department of Genetics, Washington
 UNIVERSITY, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 4 (bases 1 to 25600)
 REFERENCE
 AUTHORS Waterston, R.
 JOURNAL Direct Submission
 TITLE Submitted (25-JUN-2001) Department of Genetics, Washington
 UNIVERSITY, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 5 (bases 1 to 25600)
 REFERENCE
 AUTHORS Waterston, R.
 JOURNAL Direct Submission
 TITLE Submitted (23-MAY-2002) Department of Genetics, Washington
 UNIVERSITY, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 25600)
 REFERENCE
 AUTHORS Waterston, R.
 JOURNAL Direct Submission
 TITLE Submitted (06-NOV-2002) Department of Genetics, Washington
 UNIVERSITY, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 25600)
 REFERENCE
 AUTHORS Waterston, R.
 JOURNAL Direct Submission
 TITLE Submitted (21-NOV-2002) Department of Genetics, Washington
 UNIVERSITY, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 8 (bases 1 to 25600)
 REFERENCE
 AUTHORS Wilson, R.
 JOURNAL Direct Submission
 TITLE Submitted (15-JUN-2003) Department of Genetics, Washington
 UNIVERSITY, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 9 (bases 1 to 25600)
 REFERENCE
 AUTHORS WormBase Consortium
 CONSRTM Direct Submission
 TITLE Submitted (22-SEP-2004) Department of Genetics, Washington
 UNIVERSITY, Genome Sequencing Center, 4444 Forest Park Avenue, St.

COMMENT

Louis, MO 63110, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: submissions@watson.wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m33 subclone.

For a graphical representation of this clone sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T20F5;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is C50F2, 1000 bp overlap; the 3' clone is T21E3, 200
 bp overlap. Actual start of this clone is at base position 1 of
 T20F5; actual end is at 10147 of T21E3.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 GeneFINDER (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yujii Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFome cloning project (<http://wormfdb.dfci.harvard.edu/>),
 similarity to other proteins from Blastx analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000). Individual C. elegans Genbank submissions,
 and personal communications with C. elegans Genbank researchers,
 are predicted using the program CRMapScan-SE (Lowe, T.M. and
 Bddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

1..25600

/organism="Caenorhabditis elegans"

/mol_type="genomic DNA"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I"

/clone="T20F5"

4237..4238

/note="SL1 trans-splice acceptor; see yk34595.5"

4243..5005

/gene="pds-4"

/locus_tag="T20F5.2"

join(4243..4345,4395..4690,4741..4807,4863..5005)

/gene="pds-4"

/locus_tag="T20F5.2"

/standard_name="T20F5.2"

/note="contains similarity to Pfam domain PF00227

(Proteasome A-type and B-type); coded for by the following

C. elegans CDNAs: BE228120, CB104130, YK211c2.5,

YK333a11.5, YK34595.3, YK34595.5, YK416b12.3, YK416b12.5,

YK482b2.3, YK482b2.5, YK50798.3, YK50798.5, YK579a8.3,

YK657a8.5, YK680b1.3, YK680b1.5, YK776d09.3, YK1532c08.3,

YK1532c08.5"

/codon_start=1

/product="Proteasome beta subunit protein 4"

/protein_id="AAB37663.1"

misc_feature

gene

CDS

```
/db_xref="GI:1703615"
/db_xref="WormBase:T20F5.2"
/translation="MAGMHPVGI STEWVILLADKATPAAYGAILAADSENKEYRLCK
KLTMCTCEBQDVAFQFDMTKRNLIQVSRNGEYSPSCAHHPRRRSIABGIRSDHY
TVDVLIIGYDDEKEDAFVGSVDYLANGGQDPYLRGCGRFPYAIMDREYKDMTEA
EGLIMNKCIGEBKAFVANI PGYKVVIIIDKRYKDLDF"
complement(5177..7962)
/gene="T20F5.3"
/locus_tag="T20F5.3"
/locus_tag="T20F5.3"
complement(join(5177..5485,5531..5692,5739..5858,
5903..6033,6967..7016,7060..7154,7720..7962))
/gene="T20F5.3"
/locus_tag="T20F5.3"
/standard_name="T20F5.3"
/notes="contains similarity to pfam domain PF01765
Ribosome recycling factor; coded for by the following C.
elegans cDNAs: yk274a2.5, yk67sa7.3, yk67sa7.5,
yk1697a12.3"
/codon_start=1
/product="Hypothetical protein T20F5.3"
/protein_id="AAB37664.1"
/db_xref="GI:1703616"
/db_xref="WormBase:T20F5.3"
/translation="MMASVQPTGRAVRLEPLAIEIGEPEPSNSYDAFCDRHSAS
NMEETKTIYSLILIPSLIYFLVYNPNPKRVPEPPTVNLITPTEETLSTE
EYAAQFELAYKNDOLVLPYGSFVNGMLMTFLAVIRLAQGRAPFOKAOITSA
VYNAKKADNKKNPAPVSNLENAVVOETIKETORVEGLVEELTHFSKIDIRO
YEDVAVKLENGDKRPLSMIARVTLKSPIMINIFPDNSATKAATLQKSTLVTPQ
QEGAVLVNVPMSKEREKMASSDAKGIILNEYKKAILEITYSKDSKSSNFSRPE
AKTRREALNMKMAHOREGGLIERRKQLKQVV"
complement(8316..10689)
/gene="T20F5.4"
/locus_tag="T20F5.4"
complement(join(8316..8376,9000..9220,9537..9683,
9731..9773,9826..10047,10261..10463,10585..10689))
/gene="T20F5.4"
/locus_tag="T20F5.4"
/standard_name="T20F5.4"
/notes="contains similarity to Candida glabrata Esp2;
TR:QGVBA4; coded for by the following C. elegans cDNAs:
CK578651, CK580427, CK581704, OSTF079F11.1, OSTF079F11.1,
yk109h3.3, yk335c2.5, yk394e10.3, yk394e10.5, yk477g6.3,
yk477g6.5, yk579h10.3, yk579h10.5, yk585b9.3, yk585b9.5,
yk600d6.3, yk600d6.5, yk605f7.3, yk605f7.5, yk681g5.3,
yk681g5.5"
/codon_start=1
/product="Hypothetical protein T20F5.4"
/protein_id="AAB37665.1"
/db_xref="GI:1703617"
/db_xref="GI:1703617"
/db_xref="WormBase:T20F5.4"
/translation="MTMNHDPRLRNRVRCRSGMOELMIAKHTLLIIONLRSPIT
PESLDVOTREKKTADVLDRLSLOSVDGSSONTISTSTSTEDTKTSNRHNTY
ENSTTMYSTTRKSLSTSHSVSHSQNDVEKTSFAPVYTSKSGHGAWKTSFOR
IRPKMPSITSTNTEYLFSPKPRHPIDILTKERSAFYVLNOSKSHSIEBLSPTLS
OLSLNDEAKTPQKAEVIRWNIQKADAMFRANVYVSTEOPTKPPPLRFSRSPVR
RALPELEKTOHFAQPCIKRSPKNTKTVFAKPIAVRPSISPKVAQSTWLQSPVL"
11555..11882
/product="probable non-coding RNA"
/standard_name="standard_name=T20F5.1"
/notes="coded for by the following cDNAs: OSTF069G4_1
OSTR069G4_1"
/db_xref="WormBase:T20F5.1"
complement(13927..14664)
/gene="T20F5.5"
/locus_tag="T20F5.5"

misc_RNA

gene

CDS

CDS

gene

CDS

Query Match 44.7% Score 28.6; DB 2; Length 2560;
Best Local Similarity 72.5%; Pred No. 65;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```
14 GGTGTTCAACCGCTGGAGAGCTGTTTAGCAAAATCTGGAACGTGAAGTA 64
|||||
|db 16230 GGTGCTCAACGACCTGTAGAGATGTGAGACGAATCTGTGTGAAGCA 16180
```

```
RESULT 4
AC161379
LOCUS
DEFINITION
ACCESSION
AC161379
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNML
REFERENCE
AUTHORS
TITLE
JOURNML
REFERENCE
AUTHORS
TITLE
JOURNML
REFERENCE
AUTHORS
TITLE
JOURNML
REFERENCE
AUTHORS
TITLE
JOURNML
COMMENT
```

AC161379 189293 bp DNA linear ROD 21-JUN-2005
Mus musculus BAC clone RP23-207P12 from chromosome 17, complete
sequence.
AC161379 AC120344
AC161379.2 GI:66793614
HTG
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 189293)
Lek, S., Kozlowicz, A. and Haglund, K.
The sequence of Mus musculus BAC clone RP23-207P12
Unpublished (2001)
2 (bases 1 to 189293)
Wilson, R. K.
Direct Submission
Submitted (12-MAY-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 189293)
Wilson, R. K.
Direct Submission
Submitted (29-MAY-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 189293)
Wilson, R. K.
Direct Submission
Submitted (21-JUN-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 29, 2005 this sequence version replaced gi:63253526.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summaries Statistics
Center project name: M_BA0207P12
Drafting center: WIBR

NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e. phred quality
>=30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone, fosmid clone or direct clone walk sequence.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to obtain the consensus sequence. The
assembly was confirmed by restriction digest.
This finishing standard has slightly changed from the previous
Human standard. Specifically, standards for regions of low sequence
complexity (such as dinucleotide repeats and small unit tandem
repeats) have been relaxed. These regions are very prevalent in the
mouse genome, and the return on extended finishing efforts is
minimal.
If a sequence meets the criteria of the above statement, it needs
no comments or tags. If the criteria are not met, such as ambiguous
bases, then the region is duly annotated.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Mes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu
SOURCE INFORMATION:

The BAC library has been constructed by Kazuhiro Osegawa and Minako Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

FEATURES

Location/Qualifiers
1..189293
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-207P12"
/clone_1ib="RP23-23"

ORIGIN

Query Match 44.7%; Score 28.6; DB 9; Length 189293;
Best Local Similarity 67.8%; Pred. No. 69;
Matches 40; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

3 TCCGGAAGAGTGTCAAGCCGAGAGCTGTTAGCAAAATCTGAACTGAA 61
185149 TCCAGAGATGCTGATGACCCGACAGAACTTACAGAGAAATCAAGACAGCA 185207

RESULT 5
AC154577/c 201463 bp DNA linear HTG 30-DEC-2004
LOCUS Mus musculus chromosome 17 clone RP23-375G9, WORKING DRAFT
DEFINITION
SEQUENCE: 10 unordered pieces.
AC154577.1 GI:56900203
HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201463)
Wilson, R.K.
The sequence of Mus musculus clone
2 (bases 1 to 201463)
Unpublished
Wilson, R.K.
Direct Submission
Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Dec 30, 2004 this sequence version replaced gi:45356321.

Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0375G09
Drafting center: WIBR

Summary Statistics

Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990119
Consensus quality: 196204 bases at least Q40
Consensus quality: 197587 bases at least Q30
Consensus quality: 198347 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1271: contig of 1271 bp in length
* 1272 1371: gap of unknown length
* 1372 2636: contig of 1265 bp in length
* 2637 2736: gap of unknown length
* 2737 4059: contig of 1323 bp in length
* 4060 4159: gap of unknown length
* 4160 6162: contig of 2003 bp in length
* 6163 6262: gap of unknown length
* 6263 8067: contig of 1805 bp in length
* 8068 9486: gap of unknown length
* 9487 9586: contig of 1319 bp in length
* 9587 9877: gap of unknown length
* 9878 43196: contig of 33610 bp in length
* 43197 43296: gap of unknown length
* 43297 78014: contig of 34718 bp in length
* 78015 78115: gap of unknown length
* 78116 134599: contig of 56485 bp in length
* 134600 134699: gap of unknown length
* 134700 201463: contig of 66764 bp in length.

FEATURES

source

1..201463
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-375G9"

misc_feature

1..1271
/note="assembly_name:Contig10"

gap

1272..1371
/estimated_length=unknown

misc_feature

1372..2636
/note="assembly_name:Contig11"

gap

2637..2736
/estimated_length=unknown

misc_feature

2737..4059
/note="assembly_name:Contig33"

gap

4060..4159
/estimated_length=unknown

misc_feature

4160..6162
/note="assembly_name:Contig34"

gap

6163..6262
/estimated_length=unknown

misc_feature

6263..8067
/note="assembly_name:Contig35"

gap

8068..8167
/estimated_length=unknown

misc_feature

8168..9486
/note="assembly_name:Contig37"

gap

9487..9586
/estimated_length=unknown

misc_feature

9587..43196
/note="assembly_name:Contig39"

gap

43197..43296
/estimated_length=unknown

misc_feature

43297..78014
/note="assembly_name:Contig40"

gap

78015..78114
/estimated_length=unknown

misc_feature

78115..134599
/note="assembly_name:Contig41"

gap

134600..134699
/estimated_length=unknown

misc_feature

134700..201463
/note="assembly_name:Contig42"

ORIGIN

Query Match 44.7%; Score 28.6; DB 14; Length 201463;
Best Local Similarity 67.8%; Pred. No. 70;

Oy	3	TCCGCGAAGAAAGTGTTTCACCAACCGCTTGGAGAAGCTGTATTGCCAAATCTGCGAAGCTGGA	61
Db	170676	TCCAAGAAAGATGCTGATGAACCGCCGACAGAAACTTGTGAGGAAAATAATCAAGAACAGGCA	170618
RESULT 6			
LOCUS	AC160041		
DEFINITION	Boe taurus clone CH240-80023, *** SEQUENCING IN PROGRESS ***, 28		
ACCESSION	AC160041	280474 bp	linear HTG 01-JUL-2005
VERSION	AC160041.2	GI:68300950	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
SOURCE	Boe taurus (cow)		
ORGANISM	Boe taurus		
REFERENCE	Eukaryota; Euteleostei; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.		
AUTHORS	Munzy,D,Marie, Metzker,M,Lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alebrooke,S, Amin,A, Angulano,D, Anyalbechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biewald,K, Blair,U, Blankenburg,K, Blych,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Cessar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Catroll,L, De Anda,C, Deedrich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Georgegeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolyet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kover,C, Kovis,C, Kraft,C,L, Lebow,H, Levay,J, Lewis,L, Li,Z, Liu,J, Liu,Y, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,Z, Lorenshewa,L, Louisedge,H, Lozado,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartine,M, Mahmood,M, Malloy,K, Mangum,B, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Manthey,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nakervicius,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwokilemeh,O, Okwomodu,G, Olarinpinlaagun,A, Pal,S, Parks,K, Paternak,S, Paul,H, Perez,A, Perez,L, Plankoch,C, Plopper,F, Polndexter,A, Popovic,D, Primm,E, Pu,L,-L, Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Sheely,J, Shivratbeyn,A, Sisson,I, Sitter,C,D, Smas,D, Sneed,A, Sodergren,E, Song,X,-Z, Sorelle,R, Soes,J, Steinfelt,M, Strong,R, Sutton,A, Svatek,A, Tabors,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umanil,K, Valas,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Wilson,R, Wleczyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Zhao,S, Dunn,D, von Neiderhausern,A, Weiser,R, Smith,D,R, Holt,R.A, Smith,H.O., Weinrock,G, and Gibbs,R.A.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 280474)		

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Worley, K.C.
 Direct Submission
 Submitted (17-APR-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 280474)
 Cow Genome Sequencing Consortium.
 Direct Submission
 Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 22, 2005 this sequence version replaced gi:62718973.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: FPDO
 Center clone name: CH240-80023
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 244601 bases at least Q40
 Consensus quality: 246532 bases at least Q30
 Consensus quality: 248254 bases at least Q20
 Estimated insert size: 247370; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	28679:	contig of 28679 bp in length
*	28680	28729: gap of 50 bp
*	28730	35701: contig of 6972 bp in length
*	35702	35752: gap of 50 bp
*	35752	43864: contig of 8113 bp in length
*	43865	43914: gap of 50 bp
*	43915	56380: contig of 12466 bp in length
*	56381	56430: gap of 50 bp
*	56431	57987: contig of 1557 bp in length
*	57988	58037: gap of 50 bp
*	58038	76238: contig of 18201 bp in length
*	76239	76288: gap of 50 bp
*	76289	82247: contig of 5959 bp in length
*	82248	82297: gap of 50 bp
*	82298	112455: contig of 30158 bp in length
*	112456	112505: gap of 50 bp
*	112506	113989: contig of 1484 bp in length
*	113990	141935: gap of 27946 bp
*	141936	146586: contig of 4651 bp in length
*	146587	146636: gap of 50 bp
*	146637	159585: contig of 12949 bp in length
*	159586	159635: gap of 50 bp
*	159636	169292: contig of 9657 bp in length


```
* 52955 55025: contig of 2071 bp in length
* 55026 55040: gap of unknown length
* 55041 57657: contig of 2617 bp in length
* 57658 57672: gap of unknown length
* 57673 60530: contig of 2858 bp in length
* 60531 60546: gap of unknown length
* 60546 64221: contig of 3676 bp in length
* 64222 64236: gap of unknown length
* 64237 66744: contig of 2508 bp in length
* 66745 66759: gap of unknown length
* 66760 69537: contig of 2778 bp in length
* 69538 69552: gap of unknown length
* 69553 71778: contig of 2226 bp in length
* 71779 71793: gap of unknown length
* 71794 73966: contig of 2173 bp in length
* 73967 73981: gap of unknown length
* 73982 75981: contig of 1985 bp in length
* 75982 78741: contig of 2760 bp in length
* 78742 78756: gap of unknown length
* 78757 81358: contig of 2602 bp in length
* 81359 81373: gap of unknown length
* 81374 84070: contig of 2657 bp in length
* 84071 84085: gap of unknown length
* 84086 86914: contig of 2829 bp in length
* 86915 86929: gap of unknown length
* 86930 89248: contig of 2319 bp in length
* 89249 89263: gap of unknown length
* 89264 91256: contig of 1993 bp in length
* 91257 91271: gap of unknown length
* 91272 95588: contig of 4317 bp in length
* 95589 95603: gap of unknown length
* 95604 97920: contig of 2317 bp in length
* 97921 97935: gap of unknown length
* 97936 100805: contig of 2870 bp in length
* 100806 100820: gap of unknown length
* 100821 104770: contig of 3950 bp in length
* 104771 104785: gap of unknown length
* 104786 107676: contig of 2891 bp in length
* 107677 107691: gap of unknown length
* 107692 110889: contig of 3198 bp in length
* 110890 110904: gap of unknown length
* 110905 113390: contig of 2486 bp in length
* 113391 113405: gap of unknown length
* 113406 116553: contig of 3148 bp in length
* 116554 116568: gap of unknown length
* 116569 119084: contig of 2516 bp in length
* 119085 119099: gap of unknown length
* 119100 122456: contig of 3357 bp in length
* 122457 122471: gap of unknown length
* 122472 125606: contig of 3135 bp in length
* 125607 125621: gap of unknown length
* 125622 128558: contig of 2937 bp in length
* 128559 128573: gap of unknown length
* 128574 132627: contig of 4054 bp in length
* 132628 132642: gap of unknown length
* 132643 135451: contig of 2809 bp in length
* 135452 135466: gap of unknown length
* 135467 138584: contig of 3118 bp in length
* 138585 138599: gap of unknown length
* 138600 141494: contig of 2895 bp in length
* 141495 141509: gap of unknown length
* 141510 145190: contig of 3681 bp in length
* 145191 145205: gap of unknown length
* 145206 148105: contig of 2900 bp in length
* 148106 148120: gap of unknown length
* 148121 152192: contig of 4057 bp in length
* 152178 152192: gap of unknown length
* 152193 155249: contig of 3057 bp in length
* 155250 155264: gap of unknown length
* 155265 159702: contig of 4438 bp in length
* 159703 159717: gap of unknown length
* 159718 163856: contig of 4139 bp in length
```

```
* 163857 163871: gap of unknown length
* 163872 167346: contig of 3475 bp in length
* 167347 167361: gap of unknown length
* 167362 172779: contig of 5418 bp in length
* 172780 172794: gap of unknown length
* 172795 176577: contig of 3783 bp in length
* 176578 176592: gap of unknown length
* 176593 180517: contig of 3925 bp in length
* 180518 180532: gap of unknown length
* 180533 185863: contig of 5331 bp in length
* 185864 185878: gap of unknown length
* 185879 190385: contig of 4507 bp in length
* 190386 190400: gap of unknown length
* 190401 195486: contig of 5086 bp in length
* 195487 195501: gap of unknown length
* 195502 200348: contig of 4847 bp in length
* 200349 200363: gap of unknown length
* 200364 205613: contig of 5250 bp in length
* 205614 205628: gap of unknown length
* 205629 210848: contig of 5220 bp in length
* 210849 215971: contig of 5108 bp in length
* 215972 215986: gap of unknown length
* 215987 221028: contig of 5042 bp in length
* 221029 221043: gap of unknown length
* 221044 227181: contig of 6138 bp in length
* 227182 227196: gap of unknown length
* 227197 232414: contig of 5218 bp in length
* 232415 232429: gap of unknown length
* 232430 239757: contig of 7328 bp in length
* 239758 239772: gap of unknown length
* 239773 246849: contig of 7077 bp in length
* 246850 246864: gap of unknown length
* 246865 254487: contig of 7623 bp in length
* 254488 254502: gap of unknown length
* 254503 258238: contig of 3736 bp in length
* 258239 258253: gap of unknown length
* 258254 263882: contig of 5629 bp in length
* 263883 263897: gap of unknown length
* 263898 270762: contig of 6865 bp in length
* 270763 270777: gap of unknown length
* 270778 278318: contig of 7541 bp in length
* 278319 278333: gap of unknown length
* 278334 286621: contig of 8288 bp in length
* 286622 286636: gap of unknown length
* 286637 294136: contig of 7500 bp in length.

FEATURES
    source
        1. .294136
            /organism="Caenorhabditis elegans"
            /mol_type="genomic DNA"
            /db_xref="taxon:6239"
            /clone="Y74A11X"
        2237. .2251
            /estimated_length=unknown
            /estimated_length=4430. .4444
            /estimated_length=unknown

    gap
        2237. .2251
        4430. .4444

    gap
        2237. .2251
        4430. .4444

Query Match 44.7% Score 28.6; DB 14; Length 294136;
Best Local Similarity 72.5% Pred. No. 71;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 14 GGTGTTCAACGCGCTGGAGAGCTGTTAGCAAAATCTGGAATCTGGAAGTA 64
DB 148891 GGTGCTCAACCACTGTRGAGAGATGTGAGCGAAGATCTGTGTGGAAGCA 148941

RESULT 8
AC164240/c 172866 bp DNA linear HTG 01-JUL-2005
AC164240/c Bob taurus clone CH240-138D7, *** SEQUENCING IN PROGRESS ***, 13
DEFINITION
AC164240
AC164240
AC164240
AC164240.2 GI:68301068
ACCESSION
VERSION
```


gap 96308..96357
/estimated_length=50
gap 117108..117324
/estimated_length=217
gap 118342..118391
/estimated_length=50
gap 151211..151260
/estimated_length=50
gap 163475..163524
/estimated_length=50
gap 166740..166839
/estimated_length=unknown
gap 168329..168428
/estimated_length=unknown
gap 169602..169701
/estimated_length=unknown
gap 171294..171393
/estimated_length=unknown

Query Match 44.4%; Score 28.4; DB 14; length 172866;
Best Local Similarity 70.4%; Pred. No. 81;
Matches 38; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 10 AGAAGGTCTTCAACGCTGAGAGCTGTTAGCAAAATCTGGAAGTGAAGT 63
Db 146110 ACAAGGTCTTCAACGCTGTTAGCAAAATCTGGAAGTGAAGT 146057

RESULT 9
AC149957
LOCUS
DEFINITION Strongylocentrotus purpuratus clone R3-32B13, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC149957 145870 bp DNA linear HTG 25-JUN-2004

AC149957.1 GI:49217129
HTG; HTGS; PHASE1; HTGS; DRAFT.
Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinozoa; Echinodermata; Echinodermata; Echinodermata;
Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 145870)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Albrooks,S.L., Amaralungu,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,B., Brown,M., Bryant,N.P.,
Bunay,C., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Dayla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
Douthett,K.J., Drepper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotte,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,
Hornis,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Krtovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Liu,W.,
Lodado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Majumdar,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,N., Neal,D., Newton,J., Newton,N., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,
Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: SPCO
Center clone name: R3-32B13
Summary Statistics
Sequencing vector: plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 134633 bases at least Q40
Consensus quality: 136701 bases at least Q30
Consensus quality: 138522 bases at least Q20
Estimated insert size: 140142; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

NOT: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 13 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 2075: contig of 2075 bp in length
2076 2173: gap of unknown length
2176 3639: contig of 1464 bp in length
3640 3739: gap of unknown length
3740 4912: contig of 1173 bp in length
4913 5012: gap of unknown length
5013 6092: contig of 1080 bp in length
6093 6192: gap of unknown length
6193 9788: contig of 3596 bp in length
9789 9888: gap of unknown length
9889 11691: contig of 1803 bp in length
11692 11791: gap of unknown length
11792 22015: contig of 10124 bp in length
22016 31447: gap of unknown length
31448 31547: gap of 9432 bp in length
31548 46708: contig of 15161 bp in length
46709 46808: gap of unknown length
46809 59760: contig of 12952 bp in length
59761 59860: gap of unknown length
59861 84333: contig of 24473 bp in length
84334 84434: gap of unknown length
84435 112037: contig of 27604 bp in length
112038 112138: gap of unknown length
112139 145870: contig of 33733 bp in length.
Location/Qualifiers
1. 145870

```

/organism="Strongyloides purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="R3-32B13"
2076.2175
/estimated_length=unknown
3640.1739
/estimated_length=unknown
4913.5012
/estimated_length=unknown
6093.6192
/estimated_length=unknown
9789.9888
/estimated_length=unknown
11692.11791
/estimated_length=unknown
21916.22015
/estimated_length=unknown
31448.31547
/estimated_length=unknown
46709.46808
/estimated_length=unknown
59761.59860
/estimated_length=unknown
84334.84433
/estimated_length=unknown
112038.112137
/estimated_length=unknown

```

ORIGIN

```

Query Match      44.1%; Score 28.2; DB 14; Length 145870;
Best Local Similarity 68.4%; Pred. No. 95;
Matches 39; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

```

QY      6 GCGAAGAGGTGTTCAACGCGCTGAGAGCTTTAGCAAAATCTGAACCTGGAAG 62
Db      128556 GCCAAGCGAGTTGACGCAAGAAAGTGAAGCTAATGAATGAGTGAACGGAAG 128612

```

```

RESULT 10
AC155077/c      229461 bp      DNA      linear      HTG 01-JUL-2005
LOCUS          Bos taurus clone CH240-35K10.*** SEQUENCING IN PROGRESS ***, 22
DEFINITION    unordered pieces.
ACCESSION     AC155077
VERSION       AC155077.2 GI:68265207
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE        Bos taurus (cow)
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 229461)
Wuzny,D.,Marle, Metzger,M.,Lee, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blewett,K., Blair,J., Blankenburg,K., Blych,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavaletto,I., Cesar,H., Center,A.,
Chacho,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falle,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Gante,R., Garcia,A., Garner,T., Garza,W.,
Gebreyes,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunsberg,P., Healand,W., Hamli,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huily,S., Hume,J., Idlebird,D., Jackson,A.,

```

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louiseged,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Mallory,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,B., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,B., Milosavljevic,A., Miner,G., Minja,E., Montemayor,D., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakoelameh,O., Okwunonu,G., Olarnpungoon,A., Pal,S., Parke,K., Paeternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,B., Pu,L., Pu,M., Quiróz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S., Sanders,W., Savery,G., Scherer,S., Scott,G., Shattman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tinney,A., Trejos,Z., Umami,K., Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wleczkyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G., and Gibbs,R.A.

REFERENCE

AUTHORS

TITLE

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 229461)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 28, 2005 this sequence version replaced gi:57334865.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: FC1Z
Center clone name: CH240-35K10

----- Summary Statistics -----
Assembly program: Atlas 3.0;
Consensus quality: 218943 bases at least Q40
Consensus quality: 221244 bases at least Q30
Consensus quality: 223215 bases at least Q20
Estimated insert size: 222266; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

```
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

*	1	42348:	contig of 42348 bp in length
*		42349	gap of 50 bp
*		42399	50935: contig of 8537 bp in length
*		50936	50985: gap of 50 bp
*		50986	63009: contig of 12024 bp in length
*		63010	63644: gap of 635 bp
*		63645	69641: contig of 5997 bp in length
*		69642	69849: gap of 208 bp
*		69850	79590: contig of 9741 bp in length
*		79591	80666: gap of 1076 bp
*		80667	83794: contig of 3128 bp in length
*		83795	83894: gap of unknown length
*		83895	86322: contig of 2428 bp in length
*		86323	86601: gap of 279 bp
*		86602	88815: contig of 2214 bp in length
*		88816	88915: gap of unknown length
*		88916	97355: contig of 8440 bp in length
*		97356	97405: gap of 50 bp
*		97406	137450: contig of 40045 bp in length
*		137451	137550: gap of unknown length
*		137551	144447: contig of 6897 bp in length
*		144448	144497: gap of 50 bp
*		14498	200363: contig of 35866 bp in length
*		200364	200413: gap of 50 bp
*		200414	202340: contig of 1927 bp in length
*		202341	202390: gap of 50 bp
*		202391	204245: contig of 1855 bp in length
*		204246	204345: gap of unknown length
*		204346	205570: contig of 1225 bp in length
*		205571	205670: gap of unknown length
*		205671	206913: contig of 1243 bp in length
*		206914	207013: gap of unknown length
*		207014	208085: contig of 1072 bp in length
*		208086	208185: gap of unknown length
*		208186	209217: contig of 1032 bp in length
*		209218	209317: gap of unknown length
*		209318	211040: contig of 1723 bp in length
*		211041	211140: gap of unknown length
*		211141	212172: contig of 1032 bp in length
*		212173	212372: gap of unknown length
*		212373	213321: contig of 1049 bp in length
*		213322	213421: gap of unknown length
*		213422	229461: contig of 16040 bp in length.

```
Location/Qualifiers
1..229461
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="caxon:9913"
/clone="CH240-35K10"
42349..42398
/estimated_length=50
50936..50985
/estimated_length=50
63010..63644
/estimated_length=635
69642..69849
/estimated_length=208
79591..80666
/estimated_length=1076
83795..83894
/estimated_length=unknown
86323..86601
/estimated_length=279
88816..88915
```

[illegible]

JOURNAL

COMMENT

Submitted (30-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Aug 16, 2002 this sequence version replaced gi:22138709.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 Summary Statistics
 Center project name: H_NH0711C24

NOTICE:

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, B., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.edu>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC069314 and AC010138.

Location/Qualifiers

1. 196622

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/clone="RP11-711C24"

/clone_11b="RPCT-11"

11079..11345

/note="CpG island (%GC=60.7, o/e=0.67, #CpGs=18)"

34311..34561

/note="CpG island (%GC=57.4, o/e=1.17, #CpGs=11)"

35283..35592

/note="CpG island (%GC=73.5, o/e=0.70, #CpGs=31)"

98867..130225

/gene="CD28"

join(98867..99140,119025..119381,122040..122164,127176..127304)

/gene="CD28"

/note="Homo sapiens CD28 antigen (Tp44) (CD28), mRNA;"

H_NH0711C24.1

This gene was based on gi(5453610)"

/product="unknown"

/codon_start=1

/protein_id="AA24123.1"

/db_xref="gi:52988736"

/translation="MRLILALNLPFSIOVTGKILVKOSPMILVADNANLSCSKSY

NLPSRFPRASLHKGLDSAVECVVGYNSQQLQVYKSTGFNCGLKGNESVTRYLQNL

misc_feature

171163..171427
 /note="CpG island (%GC=62.3, o/e=0.80, #CpGs=20)"
 177943..178329
 /note="CpG island (%GC=73.1, o/e=0.60, #CpGs=34)"

ORIGIN

Query Match 43.4%; Score 27.8; DB 8; Length 196622;
 Best Local Similarity 69.1%; Pred. No. 1.3e+02;
 Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

1 GATCCGCGAGAGGTGTTCAACGCCCTGAGAGCTTTAGCAAAATCTGAA 55
 Db 157692 GAACTGAGAGCTGAGAGCAAAATCTGAGAGAGCTGAGAGCAAAATCTGGA 157638

RESULT 12

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

HSU78045 81826 bp DNA linear PRI 21-SEP-2001
 Human collagenase and stromelysin genes, complete cds, and
 metalloelastase gene, partial cds.
 U78045
 U78045.1 GI:1688257
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 1 (bases 4225 to 11906)
 Goldberg, G.I., Wilhelm, S.M., Kronberger, A., Bauer, E.A., Grant, G.A.
 and Eisen, A.Z.
 Human fibroblast collagenase. Complete primary structure and
 homology to an oncogene transformation-induced rat protein
 J. Biol. Chem. 261 (14), 6600-6605 (1986)
 3009463
 2 (bases 49921 to 57338)
 Saus, J., Quinones, S., Ocani, Y., Nagase, H., Harris, E.D. Jr. and
 Kurkinen, M.
 The complete primary structure of human matrix metalloproteinase-3.
 Identity with stromelysin
 J. Biol. Chem. 263 (14), 6742-6745 (1988)
 3360803
 3 (bases 1 to 81826)
 Shapiro, S.D., Kobayashi, D.K. and Ley, T.J.
 Cloning and characterization of a unique elastolytic
 metalloproteinase produced by human alveolar macrophages
 J. Biol. Chem. 268 (32), 23824-23829 (1993)
 8286919
 4 (bases 1 to 81826)
 Borden, P. and Heller, R.A.
 Transcriptional control of matrix metalloproteinases and the tissue
 inhibitors of matrix metalloproteinases
 Crit. Rev. Eukaryot. Gene Expr. 7 (1-2), 159-178 (1997)
 9034720
 5 (bases 1 to 81826)
 Lin, D., Duncan, M., Allen, E., Araujo, R., Aparicio, A., Chai, A.,
 Chung, E., Davis, K., Federspiel, N., Hyman, R., Kaiman, S., Komp, C.,
 Kurd, O., Lashkari, D., Lew, H., Namath, A., Oefner, P., Roberts, D.,
 Heller, R. and Davis, R.W.
 Three Matrix Metalloproteinases on 81kb of P1 insert
 Unpublished
 6 (bases 1 to 81826)
 Lin, D.
 Direct Submission
 Submitted (12-NOV-1996) Biochemistry, Stanford University, 855
 California Ave, Palo Alto, CA 94304
 Location/Qualifiers
 1. 81826
 /organism="Homo sapiens"
 /mol_type="genomic DNA"

CDS /db_xref="taxon:9606"
complement(join(4225..4334,4516..4619,5146..5308,
6418..6551,8988..9105,9266..9421,10478..10603,
10828..10976,11070..11314,11801..11905))
/function="degrades fibrillar collagen"
/note="similar to human skin collagenase encoded by
Genbank Accession Number M1509"
/codon_start=1
/product="collagenase"
/protein_id="AAB36941.1"
/db_xref="GI:1688258"
/translation="MHSFPLLLFMGVSHSPATLTQEDVDLVQKYLEKYNL
KNDGQVRRNSGVEEKLKOMDEFGLKTYGKDAETLKVMKPCGVDVQAVL
TEGNPRMEOHTLYRIENYTPDLPAVDVHAIKXAPOLMSNTPLTFKVSQGDIM
ISFVRGDRNSPDPGPGNLAHAPGPGIGADHPDEDMTPREYNLHVAAL
ELGSLGSHSTIDICALMPTSTSGDQVLAQDDIDGICAIYGSQNVQPIGQTPK
ACDSKLTPDAITTTIRGEVWFEDRPMNTNPEVEENLTVSEVPEQLPGLAEAYER
ADREVRFGKNTKWAQGVNLYHGPDIYSFGFPRTVKHIDALALEENTGKTYF
VANKWRYDEYKSMDCGPKMIADHDFGIGHKVDVAFMKDGFYFFHGTQYKEDPK
TKRILTLQKANSWFCRKN"

misc_difference 9056
/note="when compared to Genbank Accession Number X54925"
/replace="t"
complement(join(11970..11973,12037..12045,12054..12063,
12208..12217))
/note="collagenase promoter"

CDS /citation=(4)
complement(join(49921..50021,51093..51196,52346..52505,
52905..53038,53903..54047,54224..54388,55958..56083,
56222..56370,56463..56707,57234..57338))
/function="degrades matrix procollagens"
/note="similar to human matrix metalloproteinase-3 encoded
by Genbank Accession Number J03209"
/codon_start=1
/product="stromelysin"
/protein_id="AAB36942.1"
/db_xref="GI:1688259"
/translation="MKSPLILLGVAVCAVPLDGAARGDTSNMLVOKLYENYLD
EKDVQVPRRDSGVYKIRMQKFLGELVTKLDSPTLEVMKPRGCGVDVGHFRF
FPGIKRKRKTLTRYIVNTPDLPDADSAVERKALKWEETVPLTFKRLVEGDIM
ISFVAREHGDPEYPPGPGNLAHAPGPGINGDAHPDDQMTDITGTLFLVAAL
EIGSLGFSHANTALMYPLVHSLTDLRFSLQSDINGISQLYGPPDPSEPLV
TEYPPGPGPANCPLSPFADSLTREGILFLFKRHFMRSLRLSELHLISFEP
SLSPGVDAAYEVTSKDLVPIPKGNPFMRIGNEVAGVPGHITGTPPTKDAAL
SDKRNKTYFVEDKYMRPDEKRNMEGFPQIADPEPGIDSKIDANVEERGFTYF
TGSSQLEFDPNKKVTHLTKSNWLCN"

misc_difference 56526
/note="when compared to Genbank Accession Number X05232"
/replace="a"
56679
/note="when compared to Genbank Accession Number X05232"
/replace="c"
CDS /complement(join(76889..76989,77935..78041,79565..79724,
80104..80237,81059..81182,81697..81793))
/function="degrades elastin fibers"
/note="MTE: similar to human metalloelastase precursor
(matrix metalloproteinase-12), Swiss-Prot Accession Number
P39900"
/codon_start=1
/product="metalloelastase"
/protein_id="AAB36943.1"
/db_xref="GI:1688260"
/translation="DPKAVMPTKYKVIDNTRLSADDIRGQISYGPKNQRLPNP
DNXPALCDPNLSFPAVTVGNKIFPKDRFVWLVSERPKTSVNLISLWPLPSGI
EAAVEIARNQVLEFKDKYWLISNLRPEPNPKSHSGPFPKIKIDAVFNFRY
RTYFVDQYMRVDRQMDPGPKLITKNGQIGPKIDAVFSKNKYVYFQGSNQ
FEYDPLQRIKTLKSNWFCG"

misc_difference 77945
/note="when compared to Genbank Accession Number L23808"
/replace="a"
misc_difference 81755
/note="when compared to Genbank Accession Number L23808"
/replace="g"

misc_difference 81782
/note="when compared to Genbank Accession Number L23808"
/replace="a"

ORIGIN

Query Match
Best Local Similarity 69.8%; Pred. No. 1.9e+02;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Cy 3 TCCCGAAGAGCTTCAACCGCTGAGAGCTGTTAGCAAAATCTGAA 55
Db 6519 TCCGGAGTAAAGAGGATTTGCGCATGTAAATCTGATTGAGAAATPAGCAA 6571

RESULT 13
AP000922/c

LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-750P5 map 11q22, WORKING
DRAFT SEQUENCE, 35 unordered pieces.
AP000922
AP000922.2 GI:8119063
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 173359)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 173,359 genomic DNA of 11q22
Published Only in Database (1999)
2 (bases 1 to 173359)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997738.

COMMENT
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDrafc11
Center clone name: RP11-750P5
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 149299 bases at least Q40
Consensus quality: 161098 bases at least Q30
Consensus quality: 167078 bases at least Q20
Insert size: 169959; sum-of-coverage
Quality coverage: 4.01x in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of
35 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs 'N', but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 15727 contig of 15727 bp in length
15828 27186 contig of 11359 bp in length
27287 37387 contig of 10101 bp in length
37488 47288 contig of 9801 bp in length
47389 55992 contig of 8604 bp in length

56093 63632 contig of 7540 bp in length
63733 73241 contig of 9509 bp in length
73342 80659 contig of 7318 bp in length
80760 89061 contig of 8302 bp in length
89162 94975 contig of 5814 bp in length
95076 100459 contig of 5384 bp in length
100560 106303 contig of 5744 bp in length
106404 106403 contig of 5421 bp in length
111825 111924 contig of 100 bp in length
111825 111924 contig of 100 bp in length
111925 117598 contig of 5674 bp in length
117598 123774 contig of 6076 bp in length
123774 128021 contig of 4147 bp in length
128021 128121 contig of 100 bp in length
128121 128122 contig of 100 bp in length
128122 134542 contig of 6421 bp in length
134542 134642 gap of 100 bp

63632 contig of 7540 bp in length
73241 contig of 9509 bp in length
80659 contig of 7318 bp in length
89061 contig of 8302 bp in length
94975 contig of 5814 bp in length
100459 contig of 5384 bp in length
106303 contig of 5744 bp in length
106403 contig of 5421 bp in length
111924 contig of 100 bp in length
117598 contig of 5674 bp in length
123774 contig of 6076 bp in length
128021 contig of 4147 bp in length
128121 contig of 100 bp in length
128122 contig of 100 bp in length
134542 contig of 6421 bp in length
134642 gap of 100 bp

Sequence updated (26-May-2000).
NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 15727: contig of 15727 bp in length
15728 15827: gap of 100 bp
15828 27186: contig of 11359 bp in length
27187 27286: gap of 100 bp
27287 37387: contig of 10101 bp in length
37388 37487: gap of 100 bp
37488 47288: contig of 9801 bp in length
47289 47388: gap of 100 bp
47389 55922: contig of 8604 bp in length
55923 56092: gap of 100 bp
56093 63632: contig of 7540 bp in length
63633 63732: gap of 100 bp
63733 73241: contig of 9509 bp in length
73242 73341: gap of 100 bp
73342 80659: contig of 7318 bp in length
80660 80759: gap of 100 bp
80760 89061: contig of 8302 bp in length
89062 89161: gap of 100 bp
89162 94975: contig of 5814 bp in length
94976 95076: gap of 100 bp
95077 100459: contig of 5384 bp in length
100460 100559: gap of 100 bp
100560 106303: contig of 5744 bp in length
106304 106403: gap of 100 bp
106404 111824: contig of 5421 bp in length
111825 111924: gap of 100 bp
111925 117598: contig of 5674 bp in length
117598 123774: contig of 6076 bp in length
123774 128021: contig of 4147 bp in length
128021 128121: gap of 100 bp
128121 128122: gap of 100 bp
128122 134542: contig of 6421 bp in length
134543 134642: gap of 100 bp

FEATURES

source 1. 173359
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q22"
/clone="RP11-750P5"
1. 15727
/note="assembly_fragment"
15828..27186
/note="assembly_fragment clone_end:SP6 vector_side:right"
27187..37387
/note="assembly_fragment"
37488..47288
/note="assembly_fragment"
47389..55922
/note="assembly_fragment"
56093..63632
/note="assembly_fragment"
63733..73241
/note="assembly_fragment"
73342..80659
/note="assembly_fragment clone_end:17 vector_side:right"
80760..89061
/note="assembly_fragment"
89162..94975
/note="assembly_fragment"
95076..100459
/note="assembly_fragment"
100560..106303
/note="assembly_fragment"
106404..111824
/note="assembly_fragment"
111925..117598

Query Match 42.8%; Score 27.4; DB 14; Length 173359;
Best Local Similarity 69.8%; Pred. No. 1.8e+02;

Matches	377	Conservative	0	Mismatches	16	Indels	0	Gaps	0
Qy	3	TCGCGAAGAAGGTTCACAAAGCCGTGAGAGAGCTGTTACCAAAATCTGAA	55						
Db	91412	TCGCGGTGAAGAGGATTTGTGCGCATGTAGAATCGATTAGCAAAATTAAGCA	91360						
RESULT 14									
LOCUS	AP001144/c								
DEFINITION	AP001144	180399 bp	DNA	linear	HTG 30-MAY-2000				
ACCESSION	Homo sapiens chromosome 11 clone RP11-6666	map 11q22	WORKING						
VERSION	DRAFT SEQUENCE, 23 unordered pieces.								
KEYWORDS	AP001144	GI:8118496							
SOURCE	HTG, HTGS, PHASE1, HTGS, DRAFT.								
ORGANISM	Homo sapiens (human)								
	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;								
	Homnidae; Homo.								
REFERENCE	1 (bases 1 to 180399)								
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,								
TITLE	Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.								
JOURNAL	Homo sapiens 180,399 genomic DNA of 11q22								
AUTHORS	Published Only in Database (2000)								
TITLE	2 (bases 1 to 180399)								
JOURNAL	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,								
AUTHORS	Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.								
TITLE	Direct Submission								
JOURNAL	Submitted (08-FEB-2000) Masahira Hattori, The Institute of Physical								
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);								
	Kitaseo Univ., 1-15-1 Kitaseo, Sagamihara, Kanagawa 228-8555,								
	Japan (E-mail:hattori@gsc.riken.go.jp,								
	URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,								
	Fax:81-42-778-9924)								
COMMENT	On May 31, 2000 this sequence version replaced gi:6997822.								
	----- Genome Center								
	Center: RIKEN Genomic Sciences Center (GSC)								
	Center code: RIKEN								
	Web site: http://hgp.gsc.riken.go.jp/								
	Contact: hattori@gsc.riken.go.jp								
	----- Project Information								
	Center project name: HumDrafl1								
	Center clone name: RP11-6666								
	----- Summary Statistics								
	Sequencing vector: PCR products; 100% of reads								
	Chemistry: Dye-terminator ET-amersham; 100% of reads								
	Assembly program: Phrap; version 0.99029								
	Consensus quality: 163040 bases at least Q40								
	Consensus quality: 172128 bases at least Q30								
	Consensus quality: 176430 bases at least Q20								
	Insert size: 178199; sum-of-contigs								
	Quality coverage: 4.80x in Q20 bases; sum-of-contigs								

	NOTE: This is a 'working draft' sequence. It currently consists of								
	23 contigs. The true order of the pieces is not known and their								
	order in this sequence record is arbitrary. Gaps between the								
	contigs are represented as runs 'N', but the exact sizes of the gaps								
	are unknown. This record will be updated with the finished sequence								
	as soon as it is available and the accession number will be								
	preserved								
	1	32043 contig of	32043 bp in length						
	32144	52233 contig of	20090 bp in length						
	52334	70558 contig of	18225 bp in length						
	70659	83720 contig of	13062 bp in length						
	83821	91042 contig of	7222 bp in length						
	91143	100135 contig of	8993 bp in length						
	100236	108671 contig of	8436 bp in length						
	108772	119459 contig of	10688 bp in length			</			

```

JOURNAL      encoding human transporter proteins, and uses thereof
Patent: WO 0220757-A 3 14-MAR-2002;
FEATURES
SOURCE      Applera Corporation (US)
            Location/Qualifiers
            1. 34337
              /organism="Homo sapiens"
              /mol_type="unassigned DNA"
              /db_xref="taxon:9606"
ORIGIN
Query Match      42.5%; Score 27.2; DB 6; Length 34337;
Best Local Similarity 72.9%; Pred. No. 2e+02;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY      2 ATCCGCGAAGAGGTGTTCAAAGCCCTGGAGAGAGCTGTTAGCAAAAT 49
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      21458 ATGCCCGAAGAGCTTTAAATGCCCCGAGAGAGCTCTCGGCAATTAT 21505

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 361.222 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaagggttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: gb_esc1:
2: gb_esc2:
3: gb_esc3:
4: gb_hic:
5: gb_esc4:
6: gb_esc5:
7: gb_esc6:
8: gb_esc7:
9: gb_gsa1:
10: gb_gsa2:
11: gb_gsa3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31.4	49.1	717	8	CX694562	CX694562 ydb90d11.
2	29.4	45.9	588	9	B2875603	B2875603 CH240.194
3	29.2	45.6	248	9	A2121609	A2121609 RPCI-73-4
4	28.6	44.7	299	1	AV178799	AV178799 AV178799
5	28.6	44.7	300	1	AV182829	AV182829 AV182829
6	28.6	44.7	300	1	AV200654	AV200654 AV200654
7	28.6	44.7	300	5	C30991	C30991 C30991 Yuji
8	28.6	44.7	300	5	C54714	C54714 C54714 Yuji
9	28.6	44.7	300	5	C57993	C57993 C57993 Yuji
10	28.6	44.7	360	5	C63813	C63813 C63813 Yuji
11	28.6	44.7	385	8	D27678	D27678 CELK002D3R
12	28.6	44.7	608	3	BJ794144	BJ794144 BJ794144
13	28.6	44.7	667	3	BJ134236	BJ134236 BJ134236
14	28.6	44.7	669	3	BJ814155	BJ814155 BJ814155
15	28.6	44.7	677	3	BJ156049	BJ156049 BJ156049
16	28.6	44.7	703	3	BJ150848	BJ150848 BJ150848
17	28.6	44.7	726	3	BJ784546	BJ784546 BJ784546
18	28.6	44.7	749	3	BJ793333	BJ793333 BJ793333
19	28.6	44.7	762	3	BJ796838	BJ796838 BJ796838
20	28.4	44.4	1695	10	CL949298	CL949298 OGI68B004
21	28.2	44.1	300	5	CS1360	CS1360 CS1360 Yuji
22	28.2	44.1	641	6	CD309732	CD309732 StrPu691.

C	23	28.2	44.1	813	2	BF618952	BF618952 HVSMEC000
	24	28	43.8	536	10	CL368617	CL368617 RPCI44_28
	25	27.8	43.4	479	3	BJ159227	BJ159227 BJ159227
	26	27.8	43.4	522	3	BJ160169	BJ160169 BJ160169
	27	27.6	43.1	583	4	AY068267	AY068267 Schmidt
C	28	27.6	43.1	818	7	CN215701	CN215701 29524 Sub
	29	27.6	43.1	901	3	BQ223859	BQ223859 AGENCOURT
	30	27.2	42.5	571	3	AC0479141	AC0479141 RPCI-11-2
	31	27.2	42.5	614	8	DN428215	DN428215 LIB4216-1
	32	27.2	42.5	1015	10	CD398108	CD398108 ZMBP0178
	33	27	42.2	593	3	BJ558919	BJ558919 BJ558919
	34	26.8	41.9	508	8	CL334959	CL334959 RPCI44_25
	35	26.6	41.6	374	8	N98037	N98037 2098C3 czap
	36	26.6	41.6	426	5	BU495903	BU495903 PFEStoab7
	37	26.6	41.6	464	5	CA040335	CA040335 ssalrpb51
C	38	26.6	41.6	467	6	CD882439	CD882439 Fl.106114
	39	26.6	41.6	477	11	CR875824	CR875824 Sub bctof
	40	26.6	41.6	500	1	AU087244	AU087244 AU087244
	41	26.6	41.6	501	6	CB512806	CB512806 ssalrpb54
	42	26.6	41.6	508	8	CX357218	CX357218 ssalrpb51
	43	26.6	41.6	534	8	CX356249	CX356249 ssalrpb51
	44	26.6	41.6	538	10	CZ222011	CZ222011 A1RA-aae3
C	45	26.6	41.6	576	5	CA002465	CA002465 HS07J12r

ALIGNMENTS

RESULT 1
CX694562 717 bp mRNA linear EST 19-JAN-2005
LOCUS ydb90d11.y2 Sea urchin EST Lib1. Strongylocentrotus purpuratus cDNA
DEFINITION Clone ydb90d11 5', mRNA sequence.

ACCESSION CX694562
VERSION CX694562.1 GI:57956947
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus

ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinozoa; Echinodermata; Echinodermata;
Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 717)
Coffman, J.A., Robertson, A.J., Clifton, S., Page, D., Hillier, L.,
Martin, J., Wylie, T., Dante, M., Meyer, R., Theising, B., Bowers, Y.,
Gibbons, M., Konko, I., Teagareishvili, R., Rutter, E., Kennedy, S. and
Wilson, R.

TITLE Wasnu Sea Urchin EST Project
JOURNAL Unpublished (2004)
COMMENT Contact: Dr. James A. Coffman
Wasnu Sea Urchin EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

FEATURES
source
Email: est@wustl.edu
DNA sequencing by: Washington University Genome Sequencing Center
Seq primer: -28RPrOT
High quality sequence stop: 717.
location/Qualifiers
1..717
/organism="Strongylocentrotus purpuratus"
/mol_type="mRNA"
/db_xref="taxon:7668"
/clone="ydb90d11"
/lab_host="DH108"
/clone_lib="Sea urchin EST Lib1"
/note="Vector: pCMVSPORT6.1 (Invitrogen); Site 1: NotI;
Site 2: SmaI; Arrayed normalized library of full-length
cDNAs representing blaetula stage transcriptome of the sea
urchin Strongylocentrotus purpuratus, cloned into the
vector pCMVSPORT6.1 (Invitrogen)"

ORIGIN

Query Match 49.1%; Score 31.4; DB 8; Length 717;
Best Local Similarity 71.9%; Pred. No. 7;
Matches 41; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 6 GCGAAGAGGTTCACAAAGCCTGAGAGAGCTTTAGCAAAATCTGGAAGTGAAG 62
DB 359 GCCAAGAGAGGTTCAGAGCAAAAGTGAAGCCGATAGAAATGATGAGAACTGGAAG 415

RESULT 2
B2875603 588 bp DNA linear GSS 18-MAR-2003
LOCUS CH240_194F12.TJ CHORI-240 Bos taurus genomic clone CH240_194F12,
DEFINITION genomic survey sequence.
ACCESSION B2875603
VERSION B2875603.1 GI:29103008
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 388)
Zhao,S., Shetty,J., Shatsman,S., Tsagaye,G., Geer,K.,
Shwartzbeyn,A., Gebregeorgis,E., Chen,D., Riggs,F., de Jong,P.,
Crawford,A.M. and McEwan,J.C.
Bovine BAC End Sequences from Library CHORI-240
Unpublished (2003)
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the bovine BAC library CHORI-240
(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cno.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering_information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBMC) by AgResearch Ltd., New Zealand and the
Institute of Genomic Research (TIGR), USA.
Plate: 194 row: F column: 12
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..588
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="bred: Hereford"
/db_xref="taxon:9913"
/clone="CH240_194F12"
/sex="Male"
/cell_type="Blood"
/clone_1b="CHORI-240"
/note="Vector: pTARBAcl.3; Site.1: MboI; Site.2: MboI;
Hereford bull U1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match 45.9%; Score 29.4; DB 9; Length 588;
Best Local Similarity 70.9%; Pred. No. 32;
Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 9 AAGAAGTGTTCACAAAGCCTGAGAGAGCTTTAGCAAAATCTGGAAGTGAAGT 63
DB 140 ATGAAGAGGTGCAGAGACTGCACGGCTCACTAGAAAAAAGTGAAGTGAAGT 194

RESULT 3
A2121609/c 248 bp DNA linear GSS 12-MAY-2000
LOCUS A2121609

DEFINITION RPCI-23-479G12.TJ RPCI-23 Mus musculus genomic clone
ACCESSION RPCI-23-479G12, genomic survey sequence.
VERSION A2121609
KEYWORDS A2121609.1 GI:7786685
SOURCE GSS.
ORGANISM Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 248)
Zhao,S., Nierman,W., Feldlyum,T., Malek,J., Shatsman,S.,
Akintet,B., Levins,M., McGann,S., Tsagaye,G., Geer,K., Krol,M., de
Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-479G12.TJ
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 479 row: G column: 12
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..248
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-479G12"
/sex="Female"
/lab_host="DH10B"
/clone_1b="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site.1:
EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Site
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
Query Match 45.6%; Score 29.2; DB 9; Length 248;
Best Local Similarity 69.0%; Pred. No. 32;
Matches 40; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 1 GATCGGAGAGAGGTTCACAAAGCCTGAGAGAGCTTTAGCAAAATCTGGAAGT 58
DB 71 GAGGCGAAGAGAGGTTCAGATCCCGTGGAGCTGTTCAGGCACTTAGAACTG 14

RESULT 4
AV178799/c 299 bp mRNA linear EST 22-NOV-2004
LOCUS AV178799
DEFINITION AV178799 Yui1 Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone YK57b12.3', mRNA
sequence.
ACCESSION AV178799
VERSION AV178799.2 GI:55937215
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Chromadorea; Rhabdilita;

REFERENCE
AUTHORS

Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 299)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.

Expressed genes in C.elegans
Unpublished (1999)
On Jul 21, 1999 this sequence version replaced gi:5558700.
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source

1..299
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk572b12"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_1ib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

ORIGIN

Query Match 44.7%; Score 28.6; DB 1; Length 299;
Best Local Similarity 72.5%; Pred. No. 53;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

14 GGTTTCAACGCCCTGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
274 GGTGCTCAACCACTGTAGAGATGTGAGACAGAACTCTGTGTGAAGA 224

RESULT 5
AV182829/c 300 bp mRNA linear EST 21-JUL-1999
LOCUS AV182829 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
DEFINITION embryo Caenorhabditis elegans cDNA clone YK643b6 3', mRNA sequence.
ACCESSION AV182829
VERSION AV182829.1 GI:5562730
KEYWORDS EST.

SOURCE
ORGANISM
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE
AUTHORS
1 (bases 1 to 300)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.

Expressed genes in C.elegans
Unpublished (1999)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source

1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk643b6"
/sex="hermaphrodite"
/dev_stage="embryo"

ORIGIN

/clone_1ib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

Query Match 44.7%; Score 28.6; DB 1; Length 300;
Best Local Similarity 72.5%; Pred. No. 53;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

14 GGTTTCAACGCCCTGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
269 GGTGCTCAACCACTGTAGAGATGTGAGACAGAACTCTGTGTGAAGA 219

RESULT 6
AV200654/c

LOCUS AV200654 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone YK578e6 3', mRNA sequence.
ACCESSION AV200654
VERSION AV200654.1 GI:5584425
KEYWORDS EST.

SOURCE
ORGANISM
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE
AUTHORS
1 (bases 1 to 300)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.

Expressed genes in C.elegans
Unpublished (1999)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

TITLE
JOURNAL
COMMENT

FEATURES
source

1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk578e6"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="Yuji Kohara unpublished cDNA"

ORIGIN

Query Match 44.7%; Score 28.6; DB 1; Length 300;
Best Local Similarity 72.5%; Pred. No. 53;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

14 GGTTTCAACGCCCTGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
170 GGTGCTCAACCACTGTAGAGATGTGAGACAGAACTCTGTGTGAAGA 120

RESULT 7
C30991/c

LOCUS C30991 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
DEFINITION Caenorhabditis elegans cDNA clone YK267a12 3', mRNA sequence.
ACCESSION C30991
VERSION C30991.1 GI:2362787
KEYWORDS EST.

SOURCE
ORGANISM
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics
Genome Biology Lab.
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source 1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK267a12"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

ORIGIN
Query Match 44.7%; Score 28.6; DB 5; Length 300;
Best Local Similarity 72.5%; Pred. No. 53; Mismatches 14; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 14 GGTGTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGGAATCGAAGTA 64
|||||
256 GGTGCTCAACCACTGTAGAGATGTGGAGCAATCTGTGTGGAAGGA 206
|||||

RESULT 8
LOCUS C54714 300 bp mRNA linear EST 16-SEP-1997
DEFINITION C54714 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
ACCESSION C54714
VERSION C54714.1 GI:2399315
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics
Genome Biology Lab.
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source 1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK397h12"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="Yuji Kohara unpublished cDNA"

ORIGIN
Query Match 44.7%; Score 28.6; DB 5; Length 300;
Best Local Similarity 72.5%; Pred. No. 53; Mismatches 14; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 14 GGTGTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGGAATCGAAGTA 64
|||||
251 GGTGCTCAACCACTGTAGAGATGTGGAGCAATCTGTGTGGAAGGA 201
|||||

RESULT 9
LOCUS C57993 300 bp mRNA linear EST 22-SEP-1997
DEFINITION C57993 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
ACCESSION C57993
VERSION C57993.1 GI:2416698
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics
Genome Biology Lab.
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source 1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK311e7"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="Yuji Kohara unpublished cDNA"

ORIGIN
Query Match 44.7%; Score 28.6; DB 5; Length 300;
Best Local Similarity 72.5%; Pred. No. 53; Mismatches 14; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 14 GGTGTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGGAATCGAAGTA 64
|||||
273 GGTGCTCAACCACTGTAGAGATGTGGAGCAATCTGTGTGGAAGGA 223
|||||

RESULT 10
LOCUS C63813 360 bp mRNA linear EST 22-SEP-1997
DEFINITION C63813 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
ACCESSION C63813
VERSION C63813.1 GI:2422518
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES

source

1..360
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK351h10"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="Yuji Kohara unpublished cDNA"

ORIGIN

Query Match 44.7%; Score 28.6; DB 5; Length 360;
Best Local Similarity 72.5%; Pred. No. 55;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 14 GGTTGTCACACGCTGTAGCAAAATCTGGAAGTGA 64
Db 293 GGTCCTACACCACTGTAGAGATGTGACGAGAACTCTGTGTGGAAGA 343

RESULT 11
D27678 385 bp mRNA linear EST 20-NOV-1995
LOCUS D27678.1 GI:522388
DEFINITION clone YK2d3 3', mRNA sequence.
ACCESSION D27678
VERSION D27678
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 385)
Kohara, Y., Mitsuaki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
Toward an expression map of the C.elegans genome
Unpublished (1994)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
location/Qualifiers

FEATURES

source

1..385
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK2d3"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="Yuji Kohara unpublished cDNA"

ORIGIN

Query Match 44.7%; Score 28.6; DB 8; Length 385;
Best Local Similarity 72.5%; Pred. No. 56;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 14 GGTTGTCACACGCTGTAGCAAAATCTGGAAGTGA 64
Db 319 GGTCCTACACCACTGTAGAGATGTGACGAGAACTCTGTGTGGAAGA 269

RESULT 12

BT794144/c 608 bp mRNA linear EST 25-MAY-2004
LOCUS BT794144 unpublished oligo-capped cDNA library Caenorhabditis
DEFINITION BT794144 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone YK1696f10 3', mRNA sequence.
ACCESSION BT794144
VERSION BT794144.1 GI:47672941
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 608)
Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome
Unpublished (2002)
Contact: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
location/Qualifiers

1..608
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK1696f10"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="unpublished oligo-capped cDNA library"

FEATURES

source

1..608
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK1696f10"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="unpublished oligo-capped cDNA library"

ORIGIN

Query Match 44.7%; Score 28.6; DB 3; Length 608;
Best Local Similarity 72.5%; Pred. No. 61;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 14 GGTTGTCACACGCTGTAGCAAAATCTGGAAGTGA 64
Db 314 GGTCCTACACCACTGTAGAGATGTGACGAGAACTCTGTGTGGAAGA 264

RESULT 13
BT134236 667 bp mRNA linear EST 23-JAN-2002
LOCUS BT134236 unpublished oligo-capped cDNA library, C. elegans L1 stage
DEFINITION BT134236 unpublished oligo-capped cDNA clone YK1086D09 3', mRNA sequence.
ACCESSION BT134236
VERSION BT134236
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 667)
Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome
Unpublished (2002)
Contact: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
location/Qualifiers

FEATURES

source

1..667
/organism="Caenorhabditis elegans"
/mol_type="mRNA"

```

/strain="N2"
/db_xref="taxon:6239"
/clone="yk1086b09"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"

ORIGIN
Query Match          44.7%; Score 28.6; DB 3; Length 667;
Best Local Similarity 72.5%; Pred. No. 63;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 14 GGTGTTCAAGCGCTGAGAACTGTTAGCAAAATCTGGAAGTGAAGTA 64
      |||||
DB 515 GGTGCTCAACCACTGTAGAGATGTGGACGAAATCCTGTGTGGAAGGA 465

RESULT 14
LOCUS BJB14155 669 bp mRNA linear EST 27-MAY-2004
DEFINITION BJB14155 unpublished oligo-capped cDNA library, stage L4
            Caenorhabditis elegans cDNA clone yk1487d07 3', mRNA sequence.
ACCESSION BJB14155
VERSION BJB14155.1 GI:47723975
KEYWORDS EST.
SOURCE Caenorhabditis elegans
        Caenorhabditis elegans
        Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
        Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
        1 (bases 1 to 669)
        Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
        and Sugano, S.
        A complementary view of the C. elegans genome
        Unpublished (2002)
        Contact: Tadao Shin-I
        Center For Genetic Resource Information
        National Institute of Genetics
        111 Yata, Mishima, Shizuoka 411-8540, Japan
        Tel: 81-559-81-6856
        Fax: 81-559-81-6855
        Email: tshini@genes.nig.ac.jp.
        Location/Qualifiers
            1..669
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk1487d07"
            /sex="Hermaphrodite"
            /tissue_type="whole animal"
            /dev_stage="L4"
            /clone_lib="unpublished oligo-capped cDNA library, stage
            L4"

ORIGIN
Query Match          44.7%; Score 28.6; DB 3; Length 669;
Best Local Similarity 72.5%; Pred. No. 63;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 14 GGTGTTCAAGCGCTGAGAACTGTTAGCAAAATCTGGAAGTGAAGTA 64
      |||||
DB 370 GGTGCTCAACCACTGTAGAGATGTGGACGAAATCCTGTGTGGAAGGA 320

RESULT 15
LOCUS BJB156049 677 bp mRNA linear EST 24-JAN-2002
DEFINITION BJB156049 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1349c08 3', mRNA sequence.
ACCESSION BJB156049
VERSION BJB156049.1 GI:18324034

```

```

KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
          Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
          1 (bases 1 to 677)
          Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
          and Sugano, S.
          A complementary view of the C. elegans genome
          Unpublished (2002)
          Contact: Tadao Shin-I
          Center For Genetic Resource Information
          National Institute of Genetics
          111 Yata, Mishima, Shizuoka 411-8540, Japan
          Tel: 81-559-81-6856
          Fax: 81-559-81-6855
          Email: tshini@genes.nig.ac.jp.
          Location/Qualifiers
            1..677
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk1349c08"
            /sex="hermaphrodite"
            /tissue_type="whole animal"
            /dev_stage="L1"
            /clone_lib="unpublished oligo-capped cDNA library, C.
            elegans L1 stage"

ORIGIN
Query Match          44.7%; Score 28.6; DB 3; Length 677;
Best Local Similarity 72.5%; Pred. No. 63;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 14 GGTGTTCAAGCGCTGAGAACTGTTAGCAAAATCTGGAAGTGAAGTA 64
      |||||
DB 479 GGTGCTCAACCACTGTAGAGATGTGGACGAAATCCTGTGTGGAAGGA 429

Search completed: February 27, 2006, 11:09:37
Job time : 365.222 secs

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time 40.4036 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-6

Sequence score: 64
1 gatccgcgaagaaggtgttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21.*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	64	100.0	64	13	ADV65737	ADV65737 peptide A
2	60	93.8	64	13	ADV65738	ADV65738 Oligonuc
3	27.6	43.1	1706	6	ABL90642	ABL90642 Human pol
4	27.4	42.8	201	10	ADC98284	ADC98284 WPI gene
5	27.4	42.8	1257	10	ADG980834	ADG980834 Cancer re
6	27.4	42.8	81826	10	ADL13767	ADL13767 Osteoarth
7	27.4	42.8	81826	10	ADL13767	ADL13767 Osteoarth
8	27.4	42.8	81826	10	ADL13767	ADL13767 Osteoarth
9	27.2	42.5	34337	6	ABL55857	ABL55857 Human GAB
10	27.2	42.5	34337	10	ADG88376	ADG88376 Human tra
11	26.8	41.9	78	12	ADL00404	ADL00404 Antipacte
12	26.4	41.2	1137	12	ADG51259	ADG51259 Bacterial
13	26.4	41.2	1578	11	ADM03140	ADM03140 Human CDN
14	26	40.6	2681	2	AAT68323	AAT68323 Clariol
15	26	40.6	2681	2	AAV08878	AAV08878 Clariol
16	26	40.6	2681	3	AAZ52308	AAZ52308 Clariol
17	26	40.6	2760	10	AAAD47173	AAAD47173 C. brewer
18	26	40.6	3708	10	ACF57816	ACF57816 C. brewer
19	26	40.6	4434	9	ADB10028	ADB10028 Altiococ

20	26	40.6	4434	9	ADB10026	ADB10026 Altiococ
21	26	40.6	110000	9	ADB12064_09	ADB12064_09
22	25.6	40.0	350	4	AAK56389	AAK56389 Human Imm
23	25.6	40.0	426	13	ACF90670	ACF90670 Human SIR
24	25.6	40.0	1650	12	ACH87010	ACH87010 Human gen
25	25.6	40.0	2350	4	AAH45798	AAH45798 Human cel
26	25.6	40.0	3559	6	ABL99910	ABL99910 Human sec
27	25.6	40.0	3732	6	ABL99842	ABL99842 Human sec
28	25.6	40.0	10953	5	AAS29204	AAS29204 Genomic s
29	25.6	40.0	10953	6	ABS68344	ABS68344 Human DNA
30	25.6	40.0	10953	10	ADC25466	ADC25466 Human CDN
31	25.6	40.0	98865	6	ABQ78054	ABQ78054 Human Ras
32	25.6	40.0	110000	6	ABQ69245_13	ABQ69245_13
33	25.6	40.0	110000	6	ABQ67195_3	ABQ67195_3
34	25.4	39.7	834	12	ADL03259	ADL03259 DNA encod
35	25.4	39.7	269223	4	AAK28554	AAK28554 Genomic f
36	25.2	39.4	311	4	ABK71845	ABK71845 Human foe
37	25.2	39.4	311	4	AAI52189	AAI52189 Probe #20
38	25.2	39.4	311	4	AAK46290	AAK46290 Human bon
39	25.2	39.4	311	4	AAK20221	AAK20221 Human bra
40	25.2	39.4	311	4	ABS46013	ABS46013 Human liv
41	25.2	39.4	311	6	ABS20604	ABS20604 Human gen
42	25.2	39.4	443	4	ABK59333	ABK59333 Human foe
43	25.2	39.4	443	4	AAI39134	AAI39134 Probe #78
44	25.2	39.4	443	4	AAK33347	AAK33347 Human bon
45	25.2	39.4	443	4	AAK07551	AAK07551 Human bra

ALIGNMENTS

RESULT 1
ADV65737
ID ADV65737 standard; DNA; 64 BP.

XX ADV65737;

DT 10-FEB-2005 (first entry)

DE Peptide Anal3 encoding DNA SEQ ID NO:6.

XX ds; gene expression; antibiotic; antimicrobial; fungicide; cytostatic;

XX peptide Anal3.

XX Synthetic.

PN KR2004034780-A.

XX 29-APR-2004.

XX 17-OCT-2002; 2002KR-00063379.

XX 17-OCT-2002; 2002KR-00063379.

PA (BIOL-) BIOLEADERS CORP.

PA (UYCH-) UNIV CHUNGSUN CO LTD.

PA (KOR-) KOREA RES INST BIOTSCIENCE & BIOTECHNOLOG.

PI Boe HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;

PI Park YG, Sung MH;

DR WPI; 2004-577380/56.

PT Method for surface expression of peptides p5 and anal3 using pgs bca

PT gene.

PS Example 3; SEQ ID NO 6; 25pp; Korean.

XX The invention relates to a novel method for surface expression of

XX peptides p5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs)BCA

XX gene, thereby removing a purification process of peptides p5 and Anal3,

XX and using lactic acid bacteria for the surface expression, so that

XX peptide antibiotics can be cheaply and stably mass-produced. An

CC expression vector pHELIB:pgsa-P5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide P5 encoded by the nucleotide sequence set forth
CC in ADV65735, or to the peptide Anal3 encoded by the nucleotide sequence
CC set forth in ADV65737. The present sequence encodes the peptide Anal3
CC used in the invention.
XX
SQ Sequence 64 BP, 21 A, 11 C, 19 G, 13 T, 0 U, 0 Other;
Query Match 100.0%; Score 64; DB 13; Length 64;
Best Local Similarity 100.0%; Pred. No. 6.5e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DY 1 GATCGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTACGAAATCTGGAACCTGGA 60
DB 1 GATCGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTACGAAATCTGGAACCTGGA 60
DY 61 AGTA 64
DB 61 AGTA 64
DB 61 AGTA 64
RESULT 2
ADV65738/C
ID ADV65738 standard; DNA; 64 BP.
XX
XX ADV65738;
XX
XX 10-FEB-2005 (first entry)
XX
XX Oligonucleotide of the invention SEQ ID NO:7.
XX
XX ss; gene expression; antibiotic; antimicrobial; fungicide; cytostatic.
XX
XX Synthetic.
XX
XX KR2004034780-A.
XX
XX 29-APR-2004.
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
XX (BIOL-) BIOLEADERS CORP.
XX
XX (UYCH-) UNIV CHUNSUN CO LTD.
XX
XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
XX BOO HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
XX Park YG, Sung MH;
XX
XX WPI; 2004-577380/56.
XX
XX
XX Method for surface expression of peptides p5 and anal3 using pgs bsa
XX gene.
XX
XX Example 3; SEQ ID NO 7; 25bp; Korean.
XX
XX The invention relates to a novel method for surface expression of
XX peptides P5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs) bsa
XX gene, thereby removing a purification process of peptides P5 and Anal3,
XX and using lactic acid bacteria for the surface expression, so that
XX peptide antibiotics can be cheaply and stably mass-produced. An
XX expression vector pHELIB:pgsa-P5 comprises one or more genes encoding
XX poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
XX a gene for dipolar peptide antibiotics having antimicrobial, antifungal
XX and anticancer activities, wherein the dipolar peptide antibiotic has
XX homology to the peptide P5 encoded by the nucleotide sequence set forth
XX in ADV65735, or to the peptide Anal3 encoded by the nucleotide sequence
XX set forth in ADV65737. The present sequence represents an oligonucleotide
XX used in the invention to introduce the peptide Anal3 encoding DNA into

CC expression vector pHELIB:A-Anal3.
XX
SQ Sequence 64 BP, 14 A, 19 C, 11 G, 20 T, 0 U, 0 Other;
Query Match 93.8%; Score 60; DB 13; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.7e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DY 5 CGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTACGAAATCTGGAACCTGGA 64
DB 64 CGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTACGAAATCTGGAACCTGGA 5
RESULT 3
ABL90642
ID ABL90642 standard; CDNA; 1706 BP.
XX
XX ABL90642;
XX
XX 24-MAY-2002 (first entry)
XX
XX Human polynucleotide SEQ ID NO 1204.
XX
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antifungal; hepatotropic; antidiabetic; antiinflammatory; anticancer;
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US016450.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX
XX WPI; 2002-122018/16.
XX
XX P-PSDB; ABB90233.
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX prevention of neural, immune system, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX disorders.
XX
XX Claim 4; SEQ ID NO 1204; 2081bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
XX (ABB89040-ABB90444) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pce_sequences
XX
XX Sequence 1706 BP, 527 A, 350 C, 415 G, 410 T, 0 U, 4 Other;

Query Match 43.1%; Score 27.6; DB 6; Length 1706;
Best Local Similarity 67.2%; Pred. No. 14;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 CCGGCAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTAGCAAAATCTGAA 61
DB 57 CCGGCAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTAGCAAAATCTGAA 114

RESULT 4

AD98284
ID ADC98284 standard; DNA; 201 BP.

AC ADC98284;

DT 01-JAN-2004 (first entry)

DE MMP1 gene related polymorphism marker nucleotide sequence.

XX low bone mineral density; BMD; bone damage; polymorphism; osteoporosis;
KM single nucleotide polymorphism; SNP; gene; ds; human.

OS Homo sapiens.

PN WO2003054218-A2.

PD 03-JUL-2003.

PF 19-DEC-2002; 2002WO-US040948.

PR 20-DEC-2001; 2001US-0342711P.

PR 04-NOV-2002; 2002US-0423559P.

XX (INCYTE GENOMICS INC.

PI Jones KA, Valdes A, Townley DJ, Mangion J, Galwey N, Bennett S;

PI McKay I, Schaffer A;

DR WPI; 2003-559156/52.

PT Determining whether an individual is predisposed to susceptibility to low
bone mineral density (BMD) and/or bone damage, involves identifying
polymorphisms in associated genes.

PS Claim 4; Page 215; 246pp; English.

CC The present invention describes a method of determining whether an
individual is predisposed to susceptibility to low bone mineral density
(BMD) and/or bone damage comprising identifying whether the individual
has at least one polymorphism in a polynucleotide encoding a protein,
where the polynucleotide is one of 81 200-500 nucleotide sequences (S1,
see ADC98235 to ADC98315). An agent identified in an method from the
present invention which can be used for the prevention or treatment of a
disease resulting in susceptibility to low BMD and/or bone damage is
useful in the manufacture of a medicament for use in modulating the
susceptibility to low BMD and/or bone damage. The disease associated with
low BMD and/or bone damage is osteoporosis. The present sequence is used
in the exemplification of the present invention.

CC Sequence 201 BP; 62 A; 38 C; 46 G; 54 T; 0 U; 1 Other;

QY Query Match 42.8%; Score 27.4; DB 10; Length 201;

Best Local Similarity 69.8%; Pred. No. 9.3;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3 TCCGCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTAGCAAAATCTGAA 55
DB 34 TCCGCGAAGAGGTGTTGCGCATGTAGAAATCTGATTAAGAAATTAAGCAA 86

RESULT 5
ADT50834/C
ID ADT50834 standard; DNA; 1257 BP.

XX ADT50834;
AC 13-JAN-2005 (first entry)

DE Cancer related nucleic acid sequence #138.

KM ds; gene; cyrostatic; gene therapy; vaccine; diagnosis; breast; colon;
lung; ovarian; prostate; cancer.

OS Homo sapiens.

PN WO2004092338-A2.

PD 28-OCT-2004.

PF 12-APR-2004; 2004WO-US011104.

PR 11-APR-2003; 2003US-0462399P.

PR 01-JUL-2003; 2003US-0464333P.

PA (DIAD-) DIADEXUS INC.

PI Macina RA, Turner LR, Sun Y, Tam A;

DR WPI; 2004-766851/75.

PT New cancer specific nucleic acid (CasNA) molecules, useful for
diagnosing, monitoring the presence of, or treating a patient with
breast, colon, lung, ovarian, or prostate cancer.

PS Claim 1, SEQ ID NO 138; 891pp; English.

CC The invention relates to an isolated nucleic acid molecule (I)
selectively hybridizing to, or comprising at least 95% sequence identity
to, any of the 362 nucleotide sequences fully defined in the
specification. The nucleic acid molecules and polypeptides are useful for
diagnosing, monitoring the presence of, or treating a patient with
breast, colon, lung, ovarian, or prostate cancer. This sequence
corresponds to a nucleic acid of the invention.

CC Sequence 1257 BP; 397 A; 240 C; 245 G; 375 T; 0 U; 0 Other;

QY Query Match 42.8%; Score 27.4; DB 13; Length 1257;

Best Local Similarity 69.8%; Pred. No. 15;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3 TCCGCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTAGCAAAATCTGAA 55
DB 275 TCCGCGAAGAGGTGTTGCGCATGTAGAAATCTGATTAAGAAATTAAGCAA 223

RESULT 6

ADL13767
ID ADL13767 standard; DNA; 81826 BP.

AC ADL13767;

DT 06-MAY-2004 (first entry)

DE Osteoarthritis-associated polymorphic nucleotide #299.

XX ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
joint space narrowing; osteophyte development; joint pain;
osteoarthritis; SNP; single nucleotide polymorphism.

OS Homo sapiens.

PN WO2003054166-A2.

PD 03-JUL-2003.

PF 19-DEC-2002; 2002WO-US041225.


```
XX 20-DEC-2001; 2001US-0342603P.
XX (INCY-) INCYTE GENOMICS INC.
XX Jones KA, Schafer A;
XX WPI; 2003-559141/52.
XX
XX Determining susceptibility of an individual to joint space narrowing,
XX osteophyte development and/or joint pain comprises identifying whether
XX the individual has at least one polymorphism in a polynucleotide encoding
XX a protein.
XX
XX Disclosure: SEQ ID NO 299; 297bp; English.
XX
XX The invention relates to a method of determining susceptibility of an
XX individual to joint space narrowing and/or osteophyte development and/or
XX joint pain comprising identifying whether the individual has at least one
XX polymorphism in a polynucleotide encoding at least one of the protein
XX listed in the specification. The methods, composition and agent are
XX useful for modulating the susceptibility of an individual to joint space
XX narrowing and/or osteophyte development and/or joint pain that is
XX associated with a disease, preferably osteoarthritis. The cell line and
XX the non-human animal are useful for screening for an agent for diagnosing
XX an individual having susceptibility to joint space narrowing and/or
XX osteophyte development and/or joint pain. This sequence corresponds to
XX the polynucleotide encoding a protein listed in the specification. (Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences).
XX
XX Sequence 81826 BP; 23858 A; 16501 C; 15287 G; 26165 T; 0 U; 15 Other;
XX
XX Query Match 42.8%; Score 27.4; DB 10; Length 81826;
XX Best Local Similarity 69.8%; Pred. No. 45;
XX Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
XX
XX 3 TCCCGGAGAGGAGGTTCGAACGCTGGAGAGCTGTTAGCAAAATCTGGA 55
XX 6519 TCCGGGTGAGAGGAGATTGTGCGCATGTAGATCTGATTGAGAAATAGCAA 6571
XX
XX RESULT 7
XX ADL13783
XX ID ADL13783 standard; DNA; 81826 BP.
XX AC ADL13783;
XX
XX 06-MAY-2004 (first entry)
XX
XX Osteoarthritis-associated polymorphic nucleotide #315.
XX
XX ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
XX joint space narrowing; osteophyte development; joint pain;
XX osteoarthritis; SNP; single nucleotide polymorphism.
XX
XX Homo sapiens.
XX
XX WO2003054166-A2.
XX
XX 03-JUL-2003.
XX
XX 19-DEC-2002; 2002WO-US041225.
XX
XX 20-DEC-2001; 2001US-0342603P.
XX (INCY-) INCYTE GENOMICS INC.
XX Jones KA, Schafer A;
XX WPI; 2003-559141/52.
XX
```

```
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
PT
PT Disclosure: SEQ ID NO 315; 297bp; English.
PT
PT The invention relates to a method of determining susceptibility of an
PT individual to joint space narrowing and/or osteophyte development and/or
PT joint pain comprising identifying whether the individual has at least one
PT polymorphism in a polynucleotide encoding at least one of the protein
PT listed in the specification. The methods, composition and agent are
PT useful for modulating the susceptibility of an individual to joint space
PT narrowing and/or osteophyte development and/or joint pain that is
PT associated with a disease, preferably osteoarthritis. The cell line and
PT the non-human animal are useful for screening for an agent for diagnosing
PT an individual having susceptibility to joint space narrowing and/or
PT osteophyte development and/or joint pain. This sequence corresponds to
PT the polynucleotide encoding a protein listed in the specification. (Note:
PT The sequence data for this patent did not form part of the printed
PT specification but was obtained in electronic format directly from WIPO at
PT ftp.wipo.int/pub/published_pct_sequences).
PT
PT Sequence 81826 BP; 23858 A; 16501 C; 15287 G; 26165 T; 0 U; 15 Other;
PT
PT Query Match 42.8%; Score 27.4; DB 10; Length 81826;
PT Best Local Similarity 69.8%; Pred. No. 45;
PT Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
PT
PT 3 TCCCGGAGAGGAGGTTCGAACGCTGGAGAGCTGTTAGCAAAATCTGGA 55
PT 6519 TCCGGGTGAGAGGAGATTGTGCGCATGTAGATCTGATTGAGAAATAGCAA 6571
PT
PT RESULT 8
PT ADV35121
PT ID ADV35121 standard; CDNA; 81826 BP.
PT AC ADV35121;
PT
PT 10-FEB-2005 (first entry)
PT
PT Human CDNA of an exemplary efficacy gene for BAD SeqID197.
PT
PT human; ss; multi-parameter high throughput screening; MPHTS;
PT disease signature; neuropsychiatric; neurodegenerative; schizophrenia;
PT bipolar affective disorder; BAD; autism; Parkinson's;
PT Alzheimer's disease; neuroleptic; nootropic; antianemic; antidepressant.
PT
PT Homo sapiens.
PT
PT US2003096264-A1.
PT
PT 22-MAY-2003.
PT
PT 18-JUN-2002; 2002US-00175523.
PT
PT 18-JUN-2001; 2001US-0299151P.
PT 07-SEP-2001; 2001US-0317828P.
PT 25-SEP-2001; 2001US-0325150P.
PT 14-NOV-2001; 2001US-033047P.
PT 18-JAN-2002; 2002US-0349936P.
PT 04-MAR-2002; 2002US-0361834P.
PT
PT (PSYC-) PSYCHIATRIC GENOMICS INC.
PT Altar CA, Brockman JA, Evans D, Hook D, Klimczak LJ, Iaeng P;
PT Palfreyman M, Rajan P;
PT WPI; 2004-118903/12.
PT
PT Identifying a compound that can treat disease or disorders, such as, a
PT neuropsychiatric disorder e.g., schizophrenia, or autism, comprises
```

PT determining the expression of one or more efficacy genes in a cell
PT contacted with the test compound.

Example 6; SEQ ID NO 197; 39pp; English.

CC This invention relates to a novel screening method identified as a multi-
CC parameter high throughput screening (MPHS) assay. Specifically, it
CC refers to an assay that utilizes the disease signature of a plurality of
CC specific genes associated with a particular disease, and identifies
CC differential expression between those cells taken from individuals
CC affected by that disease and those that are not affected. The present
CC invention then describes the screening of candidate pharmaceutical
CC compounds to identify those that have a potential therapeutic benefit for
CC the treatment of neuropsychiatric and neurodegenerative disorders
CC including schizophrenia, bipolar affective disorder (BAD) and autism, as
CC well as Parkinson's and Alzheimer's disease. Accordingly, the compounds
CC of this invention exhibit various activities including neuroleptic,
CC neurotropic, anti manic and antidepressant. Furthermore, the screening
CC method used in MPHS will be automated, such that a large number of test
CC compounds may be rapidly screened with a minimal amount of labour and
CC effort. This polynucleotide is a human cDNA sequence of a gene that is
CC differentially expressed in the presence of a therapeutic compound and
CC represents an exemplary efficacy gene for bipolar affective disorder,
CC given in an exemplification of the invention.

XX Sequence 81826 BP; 23858 A; 16501 C; 15287 G; 26165 T; 0 U; 15 Other;

Query Match 42.8%; Score 27.4; DB 13; Length 81826;

Best Local Similarity 69.8%; Pred. No. 45;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 3 TCCGCGAAGAGGTGTTCAAAAGCGCTGAGACCTGTTGCAAAATCTGGA 55
Db 6519 TCCGGGTAGAGGAGATTGTGCGCATGTGATCTGATGAATAATAGCAA 6571

RESULT 9
ABLS5857
ID ABL55857 standard; DNA; 34337 BP.

XX ABL55857;

XX 15-JUL-2002 (first entry)

DE Human GABA transporter protein gene.

XX Human; GABA; transporter protein; drug screening; therapeutic target;

XX gene; ds.

XX Homo sapiens.

XX Key

XX variation

XX CDS

XX exon

XX intron

XX variation

XX variation

XX variation

XX variation

XX variation

XX variation

FT /standard_name= "Single nucleotide polymorphism"
FT replace(644,G)
FT /tag= h

FT /standard_name= "Single nucleotide polymorphism"
FT replace(7538,A)
FT /tag= i

FT /standard_name= "Single nucleotide polymorphism"
FT 8040. .8197
FT /tag= j

FT /number= 2
FT 8198. .8564
FT /tag= k

FT /number= 2
FT 8547. .8673
FT /tag= l

FT /number= 3
FT 8674. .10067
FT /tag= m

FT /number= 3
FT replace(9127,G)
FT /tag= n

FT /standard_name= "Single nucleotide polymorphism"
FT replace(9256,G)
FT /tag= o

FT /standard_name= "Single nucleotide polymorphism"
FT replace(9421,G)
FT /tag= p

FT /standard_name= "Single nucleotide polymorphism"
FT 10068. .10249
FT /tag= q

FT /number= 4
FT 10250. .17500
FT /tag= r

FT /number= 4
FT replace(13248. .13250,TC)
FT /tag= s

FT /standard_name= "Single nucleotide polymorphism"
FT replace(13377,G)
FT /tag= t

FT /standard_name= "Single nucleotide polymorphism"
FT replace(13434,C)
FT /tag= u

FT /standard_name= "Single nucleotide polymorphism"
FT replace(13672,A)
FT /tag= v

FT /standard_name= "Single nucleotide polymorphism"
FT replace(14502,T)
FT /tag= w

FT /standard_name= "Single nucleotide polymorphism"
FT replace(14774. .14776,AAA/AA)
FT /tag= x

FT /standard_name= "Single nucleotide polymorphism"
FT replace(17289,A)
FT /tag= y

FT /standard_name= "Single nucleotide polymorphism"
FT 17501. .17611
FT /tag= z

FT /number= 5
FT 17612. .20639
FT /tag= aa

FT /number= 5
FT replace(17660,G)
FT /tag= ab

FT /standard_name= "Single nucleotide polymorphism"
FT replace(19702,G)
FT /tag= ac

FT /standard_name= "Single nucleotide polymorphism"
FT 20640. .20881
FT /tag= ad

FT /number= 6
FT 20882. .21173
FT /tag= ae

FT /number= 6
FT replace(6425,C)
FT /tag= g

FT exon 21174..21366
FT /*tag= af
FT /number= 7
FT replace(21263,T)
FT /*tag= ag
FT /standard_name= "single nucleotide polymorphism"
FT 21367..21397
FT /*tag= ah
FT /number= 7
FT 23298..23490
FT /*tag= ai
FT /number= 8
FT 23491..26774
FT /*tag= aj
FT /number= 8
FT replace(24502..24504,TG)
FT /*tag= ak
FT /standard_name= "single nucleotide polymorphism"
FT replace(24642,C)
FT /*tag= al
FT /standard_name= "single nucleotide polymorphism"
FT 26775..26934
FT /*tag= am
FT /number= 9
FT 26935..30367
FT /*tag= an
FT /number= 9
FT replace(28543,T)
FT /*tag= ao
FT /standard_name= "single nucleotide polymorphism"
FT replace(30303,C)
FT /*tag= ap
FT /standard_name= "single nucleotide polymorphism"
FT 30368..30530
FT /*tag= aq
FT /number= 10
FT 30531..31962
FT /*tag= ar
FT /number= 10
FT 31963..32337
FT /*tag= as
FT /number= 11
FT exon
FT US2002031800-A1.
FT 14-MAR-2002.
FT 21-DEC-2000; 2000US-00741149.
FT 05-SEP-2000; 2000US-0229529P.
FT
FT (LIZZ/) LI Z.
FT (CHAT/) CHATURVEDI K.
FT (ZHUS/) ZHU S.
FT (WOOD/) WOODAGE T.
FT (GUEG/) GUEGLER K.
FT (WEBER/) WEBSTER M.
FT (DEFR/) DI FRANCESCO V.
FT (BEAS/) BEASLEY E M.
FT
FT Li Z, Chaturvedi K, Zhu S, Woodage T, Guegler K, Webster M;
FT Di Francesco V, Beasley EM;
FT WPI: 2002-361179/39.
FT P-PSDB; ABB77168.
FT
FT New peptides and nucleic acid sequences related to the GABA transporter
FT subfamily, useful in developing diagnostic compositions, as well as in
FT drug screening, particularly as models for the development of human
FT therapeutic targets.
FT
FT Claim 2; Fig 3; 69pp; English.
FT
FT

CC The sequence encodes a novel human transporter protein of the GABA
CC transporter subfamily. The invention relates to a novel isolated human
CC peptide of the GABA transporter subfamily. The peptide and nucleic acids
CC are useful in developing human therapeutics and diagnostic compositions.
CC These are also useful in drug screening, particularly as models for the
CC development of human therapeutic targets
XX
SQ Sequence 34337 BP; 10987 A; 5399 C; 6061 G; 11588 T; 0 U; 302 Other;
Query March 42.5%; Score 27.2; DB 6; Length 34337;
Best Local Similarity 72.9%; Pred. No. 42;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 2 ATCCGCGAAGAGGTGTTCAACCGCTGAGAACTGTTAGCAAAAT 49
Db 21458 ATGCCCGAAGAGCTCTTAATAATGCCGGAAGAACTCTGGGCAATAT 21505
RESULT 10
ADG88376
ID ADG88376 standard; DNA; 34337 BP.
XX
AC ADG88376;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human transporter gene.
XX
KW Human; transporter protein; GABA transporter; therapeutic protein;
KW pharmacogenomic analysis; immune response; biological fluid;
KW foetal heart tissue; gene therapy; gene; de.
XX
OS Homo sapiens.
XX
FH Key
FH CDS
FT Location/Qualifiers
FT 1001..31067
FT /*tag= b
FT /product= "Human transporter protein"
FT 1001..1289
FT /*tag= a
FT 1290..7053
FT /*tag= c
FT 7054..7211
FT /*tag= d
FT 7212..7560
FT /*tag= e
FT 7561..7687
FT /*tag= f
FT 7688..9081
FT /*tag= g
FT 9082..9263
FT /*tag= h
FT 9264..16514
FT /*tag= i
FT 16515..16625
FT /*tag= j
FT 16626..19366
FT /*tag= k
FT 19367..19608
FT /*tag= l
FT 19609..19900
FT /*tag= m
FT 19901..20093
FT /*tag= n
FT 20094..22026
FT /*tag= o
FT 22027..22219
FT /*tag= p
FT 22220..25504
FT /*tag= q
FT 25505..25664
FT /*tag= r
FT 25665..29097
FT intron

PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 29689; 122bp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX CC promoter functional in a plant cell, where the promoter is positioned to
XX CC provide for expression of a polynucleotide encoding a polypeptide from a
XX CC microbial source. The invention also relates to a transformed plant
XX CC comprising the recombinant DNA construct and a method of producing a
XX CC transgenic plant having an improved property. The plant is a crop plant
XX CC such as maize or soybean. The method of producing a transformed plant
XX CC having an improved property comprises transforming a plant with the
XX CC recombinant DNA construct and growing the transformed plant, where the
XX CC polynucleotide or polypeptide is useful for improving plant properties.
XX CC The recombinant DNA construct is useful for producing plants with
XX CC improved plant properties, e.g. improved cold, heat or drought tolerance,
XX CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX CC increased resistance to plant disease, better growth rate by modification
XX CC of the cell cycle pathway with plant growth regulators, increased rate of
XX CC homologous recombination, modified seed oil or protein yield and/or
XX CC content, improved yield by modification of carbohydrate, nitrogen or
XX CC phosphorus use and/or uptake, by modification of photosynthesis or by
XX CC providing improved plant growth and development under at least one stress
XX CC condition, improved lignin production or improved galactomannan
XX CC production. This sequence represents a bacterial polynucleotide used in
XX CC the scope of the invention. Note: The sequence data for this patent did
XX CC not form part of the printed specification but was obtained in electronic
XX CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1137 BP; 282 A; 266 C; 278 G; 311 T; 0 U; 0 Other;
XX
Query Match 41.2%; Score 26.4; DB 13; Length 1137;
Best Local Similarity 65.0%; Pred. No. 33;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
XX
QY 1 GATCGCCGAAGAAGTGTTCGAACGCTCGAGAAGCTGTTGCAAAATCTGGAATCGA 60
DB 429 GATCCAGGAATCAGCTGTGCAACGCTCTGCAAGCCGAACATCAAGTCAGATCAGCA 488
XX
RESULT 13
ADM03140
ID ADM03140 standard; cDNA; 1578 BP.
XX
XX ADM03140;
XX
XX 20-MAY-2004 (first entry)
XX
XX Human cDNA of the invention SEQ ID NO:1825.
XX
XX ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
XX Homo sapiens.
XX
XX EP1347046-A1.
XX
XX 24-SEP-2003.
XX
XX 12-APR-2002; 2002EP-00008400.
XX
XX 22-MAR-2002; 2002JP-00137785.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
XX WPI; 2003-723556/69.
XX
XX P-PSDB; ADM05583.
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for
XX PT developing a diagnostic marker or medicines for regulating their
XX PT expression and activity, or as a target of gene therapy.
XX
XX Claim 1; SEQ ID NO 1825; 305bp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
XX CC polypeptide. A polynucleotide of the invention may have a use in gene
XX CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
XX CC as a primer for synthesizing the polynucleotide or as a probe for
XX CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
XX CC useful in gene therapy, for developing a diagnostic marker or medicines
XX CC for regulating their expression and activity, or as a target of gene
XX CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
XX CC are useful as pharmaceutical agents. The present sequence represents a
XX CC cDNA sequence of the invention.
XX
SQ Sequence 1578 BP; 505 A; 320 C; 329 G; 424 T; 0 U; 0 Other;
XX
Query Match 41.2%; Score 26.4; DB 11; Length 1578;
Best Local Similarity 69.2%; Pred. No. 36;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
XX
QY 11 GAGGTGTTCAACGCGCTGGAAGAGCTGTTCGAAATCTGGAATCTGGAAG 62
DB 546 GACATGTTCAGACGACCTGGAGGAAGCATTTGATGATGATTCCTGAACCTGAAG 597
XX
RESULT 14
AAT68323/C
ID AAT68323 standard; cDNA; 2681 BP.
XX
XX AAT68323;
XX
XX 09-AUG-1997 (first entry)
XX
XX Clarkia breweri S-linalool synthase cDNA.
XX
XX DE Clarkia breweri S-linalool synthase cDNA.
XX
XX KM S-linalool synthase; scent; fragrance; aroma; flavour; transgenic plant;
XX
XX ss.
XX
XX Clarkia breweri.
XX
XX OS
XX
XX Key Location/Qualifiers
XX
XX FT CDS 28..2640
XX
XX FT /*tag= a
XX
XX MO9715584-A2.
XX
XX PN 01-MAY-1997.
XX
XX PD
XX
XX 15-OCT-1996; 96WO-US016482.
XX
XX PF 12-OCT-1995; 95US-0005146P.
XX
XX PR
XX
XX (UNMI) UNIV MICHIGAN.
XX
XX PA
XX
XX Pichereky E;
XX
XX PI
XX
XX WPI; 1997-258947/23.
XX
XX DR P-PSDB; AAM17080.
XX
XX Linalool synthase from Clarkia breweri (Onograceae) - an acyclic
XX PT monoterpene, useful for enhancing the scent production or flavour of
XX PT plants.
XX

PS Claim 3; Page 73-82; 105pp; English.

CC A CDNA clone (AA17080) codes for Clarkia breweri S-linalool synthase
 CC (AA17080), an enzyme that catalyzes the conversion of geranyl-
 CC pyrophosphate to S-linalool, an acyclic monoterpene that is a major
 CC component of the plant's scent. It was obtained by PCR amplification of cDNA
 CC from petal and stigmata cDNA libraries. The clone can be used to produce
 CC S-linalool synthase in host cells, to enhance scent production in a
 CC transgenic plant such as petunia, rose, carnation etc., or to enhance the
 CC flavour of a transgenic plant such as tomato, grape and tea

SO Sequence 2681 BP; 820 A; 607 C; 532 G; 722 T; 0 U; 0 Other;

Query Match 40.6%; Score 26; DB 2; Length 2681;

Best Local Similarity 65.5%; Pred. No. 58;

Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 6 GCGAAGAGGTGTTCAACGCTGGAGAGCTGTTAGCAAAATCTGAACTGGAAGT 63
 DB 939 GAGTAGAAGGTGTCATCTCCGGAGTAGAGATTGTCACAAACCGACTTGTCAGT 882

RESULT 15

AAV08878/C

ID AAV08878 standard; DNA; 2681 BP.

AC AAV08878;

DT 19-FEB-1999 (first entry)

DE Linalool synthase coding sequence.

KW Linalool synthase; scent enhancer; ss.

OS Clarkia breweri.

Key Location/Qualifiers

FT CDS 28..2640

FT /*tag= a

PN US5849526-A.

PD 15-DEC-1998.

PF 15-OCT-1996; 96US-00732192.

PR 15-OCT-1996; 96US-00732192.

PA (UNMI) UNIV MICHIGAN.

PI Pichersky B;

DR WPI; 1999-069727/06.

DR P-PSDB; AAW73485.

PT DNA encoding Clarkia linalool synthase protein - for producing recombinant protein or transgenic plants.

PS Claim 7; Col 33-40; 30pp; English.

CC This sequence represents the nucleic acid of the invention, encoding the
 CC linalool synthase protein of Clarkia breweri. The DNA is used to produce
 CC recombinant Clarkia breweri S-linalool synthase or to produce transgenic
 CC plants that express Clarkia breweri S-linalool synthase (especially for
 CC the production of enhanced scent and taste in plants)

SO Sequence 2681 BP; 820 A; 607 C; 532 G; 722 T; 0 U; 0 Other;

Query Match 40.6%; Score 26; DB 2; Length 2681;

Best Local Similarity 65.5%; Pred. No. 58;

Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

DB 939 GAGTAGAAGGTGTCATCTCCGGAGTAGAGATTGTCACAAACCGACTTGTCAGT 882

Search completed: February 27, 2006, 07:05:28
 Job time: 44.6036 secs

THIS PAGE BLANK (015120)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceeleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28 ; Search time 70.0816 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
1 gatccgcgaagaagggtgttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	64	9 US-10-789-164-6	Sequence 6, Appl1
2	60	93.8	64	9 US-10-789-164-7	Sequence 7, Appl1
3	28.4	44.4	1659	7 US-10-437-963-86954	Sequence 86954, A
4	27.6	43.1	1706	6 US-10-264-237-1204	Sequence 1204, Ap
5	27.4	42.8	81826	5 US-10-175-523-197	Sequence 197, App
6	27.4	42.8	81826	10 US-11-099-266-197	Sequence 197, App
7	27.2	42.5	554	4 US-09-925-065A-227036	Sequence 227036, A
8	27.2	42.5	556	4 US-09-925-065A-171598	Sequence 171598, A
9	27.2	42.5	34337	4 US-09-741-149-3	Sequence 3, Appl1
10	27.2	42.5	34337	6 US-10-385-614-3	Sequence 3, Appl1
11	26.4	41.2	1137	6 US-10-369-493-29689	Sequence 29689, A
12	26.4	41.2	1578	6 US-10-108-260A-1825	Sequence 1825, Ap
13	26.4	41.2	2622	7 US-10-437-963-78386	Sequence 78386, A
14	26	40.6	118	8 US-10-425-115-150443	Sequence 150443, A
15	26	40.6	201	8 US-10-741-600-33231	Sequence 33231, A
16	26	40.6	2760	7 US-10-469-893-1	Sequence 1, Appl1
17	26	40.6	3708	7 US-10-376-931-1	Sequence 1, Appl1
18	26	40.6	4434	9 US-10-501-282-3597	Sequence 3597, Ap
19	26	40.6	4434	9 US-10-501-282-3599	Sequence 3599, Ap
20	26	40.6	103660	8 US-10-741-600-17645	Sequence 17645, A
21	26	40.6	1754382	9 US-10-501-282-6651	Sequence 6651, Ap
22	25.6	40.0	1650	6 US-10-029-386-20205	Sequence 20205, A
23	25.6	40.0	2125	6 US-10-115-831-74	Sequence 74, Appl1

24	25.6	40.0	10953	3 US-09-764-846-313	Sequence 313, App
25	25.6	40.0	10953	5 US-10-091-483-313	Sequence 313, App
26	25.6	40.0	98865	3 US-09-770-689A-3	Sequence 3, Appl1
27	25.6	40.0	98865	8 US-10-949-419-3	Sequence 3, Appl1
28	25.6	40.0	495269	7 US-10-398-221-8	Sequence 8, Appl1
29	25.6	40.0	3011208	7 US-10-398-221-2058	Sequence 2058, Ap
30	25.4	39.7	269223	7 US-10-672-787-41	Sequence 41, Appl1
31	25.2	39.4	311	3 US-09-864-761-28288	Sequence 28288, A
32	25.2	39.4	443	3 US-09-864-761-11712	Sequence 11712, A
33	25.2	39.4	580	3 US-09-864-761-11859	Sequence 15859, A
34	25.2	39.4	751	7 US-10-424-599-49013	Sequence 49013, A
35	25.2	39.4	2776	6 US-10-247-671-25	Sequence 25, Appl1
36	25.2	39.4	2911	6 US-10-439-741-25	Sequence 25, Appl1
37	25.2	39.4	3113	6 US-10-252-157-151	Sequence 151, App
38	25.2	39.4	3199	5 US-10-044-090-440	Sequence 440, App
39	25.2	39.4	3199	6 US-10-240-965-267	Sequence 267, App
40	25.2	39.4	4537	6 US-10-240-965-266	Sequence 266, App
41	25.2	39.4	4537	6 US-10-044-090-439	Sequence 439, App
42	25.2	39.4	4601	6 US-10-252-157-150	Sequence 150, App
43	25.2	39.4	4713	3 US-09-814-353-19757	Sequence 19757, A
44	25.2	39.4	4750	8 US-10-723-860-6880	Sequence 6880, Ap
45	25.2	39.4	6048	10 US-11-097-143-24461	Sequence 24461, A

ALIGNMENTS

RESULT 1
US-10-789-164-6
; Sequence 6, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Kopatentlm 1.71
; SEQ ID NO 6
; LENGTH: 64
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-6

Query Match 100.0%; Score 64; DB 9; Length 64;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAAGGTGTTCAACGCTGAGAGAGCTGTTAGCAAAATCTGGAAGTGA 60
DB 1 GATCCGGAAGAAGGTGTTCAACGCTGAGAGAGCTGTTAGCAAAATCTGGAAGTGA 60
QY 61 AGTA 64
DB 61 AGTA 64

RESULT 2
US-10-789-164-7/c
; Sequence 7, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9


```
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 7
; LENGTH: 64
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-7

Query Match          93.8%; Score 60; DB 9; Length 64;
Best Local Similarity 100.0%; Pred. No. 1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 CGCGAAGAGGTTTCAACCGCTGAGAGAGCTTTTACCAAAATCTGGAAGTGA 64
Db 64 CGCGAAGAGGTTTCAACCGCTGAGAGAGCTTTTACCAAAATCTGGAAGTGA 5

RESULT 3
US-10-437-963-86954
; Sequence 86954, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Bouharov, Andrei A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86954
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85948C.1
US-10-437-963-86954

Query Match          44.4%; Score 28.4; DB 7; Length 1659;
Best Local Similarity 66.1%; Pred. No. 8;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Oy 2 ATCCCGAAGAGGTTTCAACCGCTGAGAGAGCTTTTACCAAAATCTGGAAGTGA 61
Db 814 ATCCCGAAGAGGTTTCAACCGCTGAGAGAGAGCATGACCAATCTGCTGGTA 873

Oy 62 GT 63
Db 874 CT 875

RESULT 4
US-10-264-237-1204
; Sequence 1204, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA31PI
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2676
```

```
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 1204
; LENGTH: 1706
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-237-1204

Query Match          43.1%; Score 27.6; DB 6; Length 1706;
Best Local Similarity 67.2%; Pred. No. 16;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Oy 4 CGCGAAGAGGTTTCAACCGCTGAGAGAGCTTTTACCAAAATCTGGAAGTGA 61
Db 57 CCATGAAGAGGTTTCAACCGCTGAGAGAGCTTTTACCAAAATCTGGAAGTGA 114

RESULT 5
US-10-175-523-197
; Sequence 197, Application US/10175523
; Publication No. US2003009626A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Paley, Michael
; APPLICANT: Rajan, Priti
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/U795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 197
; LENGTH: 81826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(81826)
; OTHER INFORMATION: where n may be a or g or c or t/u, unknown, or other
US-10-175-523-197

Query Match          42.8%; Score 27.4; DB 5; Length 81826;
Best Local Similarity 69.8%; Pred. No. 65;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 3 TCCGGAAGAGGTTTCAACCGCTGAGAGAGCTTTTACCAAAATCTGGAAGTGA 55
Db 6519 TCCGGAAGAGGTTTCAACCGCTGAGAGAGCTTTTACCAAAATCTGGAAGTGA 6571

RESULT 6
US-11-099-266-197
; Sequence 197, Application US/11099266
; Publication No. US20050181433A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
```

APPLICANT: Klineczak, Leszek
APPLICANT: Laeng, Pascal
APPLICANT: Palitayman, Michael
APPLICANT: Rajan, Prithi
TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
FILE REFERENCE: 03235/1004795-USA
CURRENT APPLICATION NUMBER: US/11/099,266
PRIOR FILING DATE: 2005-04-04
PRIOR APPLICATION NUMBER: US 10/175,523
PRIOR FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/299,151
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/317,828
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 60/349,936
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/361,834
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn version 3.1
SEQ ID NO 197
LENGTH: 81826
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(81826)
OTHER INFORMATION: where n may be a or g or c or t/u, unknown, or other
US-11-099-266-197

Query Match 42.8%; Score 27.4; DB 10; Length 81826;
Best Local Similarity 69.8%; Pred. No. 65;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3 TCCGCGAAGAGGTGTTCAACCGCTGAGAGCTGTTTAGCAAAATCTGGAA 55
DB 6519 TCCGCGTAGAAGGATTGTGCGCATGTGAATCTGATTGGAATAATGACAA 6571

RESULT 7

US-09-925-065A-227036/c
Sequence 227036, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 95086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 227036
LENGTH: 554
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-227036

Query Match 42.5%; Score 27.2; DB 4; Length 554;

Best Local Similarity 64.1%; Pred. No. 15;
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGGTGTTCAACCGCTGAGAGCTGTTTAGCAAAATCTGGAACTGGA 60
DB 236 GCTCCAAAGAGATTGTGTCAGGCTCATGAGAAATCTTGAACCAAAATTCGAATTGGA 237

QY 61 AGTA 64
DB 236 GGTA 233

RESULT 8

US-09-925-065A-171598/c
Sequence 171598, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 171598
LENGTH: 556
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-171598

Query Match 42.5%; Score 27.2; DB 4; Length 556;
Best Local Similarity 64.1%; Pred. No. 15;
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGGTGTTCAACCGCTGAGAGCTGTTTAGCAAAATCTGGAACTGGA 60
DB 236 GCTCCAAAGAGATTGTGTCAGGCTCATGAGAAATCTTGAACCAAAATTCGAATTGGA 237

QY 61 AGTA 64
DB 236 GGTA 233

RESULT 9

US-09-741-149-3
Sequence 3, Application US/09741149
Patent No. US20020031800A1
GENERAL INFORMATION:
APPLICANT: Li, Zhenya et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
FILE REFERENCE: CLO00780
CURRENT APPLICATION NUMBER: US/09/741,149
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 34337
TYPE: DNA
ORGANISM: Human
US-09-741-149-3

NAME/KEY: misc.feature
LOCATION: (1)_(34337)
OTHER INFORMATION: n = A,T,C or G
US-09-741-149-3

Query Match 42.5%; Score 27.2; DB 3; Length 34337;
Best Local Similarity 72.9%; Pred. No. 58;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 2 ATCCGCGAAGAGCTGTTCAACCGCTGAGAGAGCTGTTTACCAAAAT 49
DB 21458 ATGCCCGAAGAGCTCTTAATAATGCCGAGAGAGCTCCCTGGCATTAT 21505

RESULT 10
US-10-385-614-3
; Sequence 3, Application US/10385614
; Publication No. US20030157649A1
; GENERAL INFORMATION:
; APPLICANT: Li, Zhenya et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO00780
; CURRENT APPLICATION NUMBER: US/10/385,614
; CURRENT FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 34337
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(34337)
; OTHER INFORMATION: n = A,T,C or G
US-10-385-614-3

Query Match 42.5%; Score 27.2; DB 6; Length 34337;
Best Local Similarity 72.9%; Pred. No. 58;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 2 ATCCGCGAAGAGCTGTTCAACCGCTGAGAGAGCTGTTTACCAAAAT 49
DB 21458 ATGCCCGAAGAGCTCTTAATAATGCCGAGAGAGCTCCCTGGCATTAT 21505

RESULT 11
US-10-369-493-29689
; Sequence 29689, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 29689
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-10-369-493-29689

Query Match 41.2%; Score 26.4; DB 6; Length 1137;
Best Local Similarity 65.0%; Pred. No. 37;

Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
OY 1 GATCCGCGAAGAGCTGTTCAACCGCTGAGAGAGCTGTTTACCAAAATCTGAACTGGA 60
DB 429 GATCCGAGATCAGCTGTTCACCAAGCCCTCTGCAAGCGGAACATCAAGCTCAGGATCAGA 488

RESULT 12
US-10-108-260A-1825
; Sequence 1825, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1825
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1825

Query Match 41.2%; Score 26.4; DB 6; Length 1578;
Best Local Similarity 69.2%; Pred. No. 41;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 11 GAAGGTGTTCAACCGCTGAGAGAGCTGTTTACCAAAATCTGAACTGGAAG 62
DB 546 GACGATGTTCAAGCAGCTGGAGAGCAATTGAGATGATTCCTGAACCTTGAAG 597

RESULT 13
US-10-437-963-78386
; Sequence 78386, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Mu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 78386
; LENGTH: 2622
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_78199C.1
US-10-437-963-78386

Query Match 41.2%; Score 26.4; DB 7; Length 2622;
Best Local Similarity 65.0%; Pred. No. 49;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 5 CGCGAAGAGGTGTTCAACCGCTGAGAGAGCTGTTTACCAAAATCTGAACTGGAAGTA 64
DB 359 CGTGTGAAGCTGTTCAACCAAGCTGAGAGAGCCGCTGCTTCACTGGAAGGAGATCA 418

RESULT 14
US-10-425-115-150443
; Sequence 150443, Application US/10425115

Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21 (5322)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 150443
LENGTH: 118
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_6872C.1
US-10-425-115-150443

Query Match 40.6%; Score 26; DB 8; Length 118;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6 GCGAAGAGGTGTTCAAGCCCTGAGAGAGCTTTAGCAAAATCTGGAA 55
DB 45 GCGCAGAGGTGTTCAAGCGGTTGCGCAGCGGCTGAGCGACATCTTGA 94

RESULT 15
US-10-741-600-33231/C
Sequence 33231, Application US/10741600
Publication No. US20050026169A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CL001499
CURRENT APPLICATION NUMBER: US/10/741,600
CURRENT FILING DATE: 2003-12-22
NUMBER OF SEQ ID NOS: 73997
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 33231
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-741-600-33231

Query Match 40.6%; Score 26; DB 8; Length 201;
Best Local Similarity 70.0%; Pred. No. 30;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 8 GAGGAAGCTGTTCAAGCCTGGAGAGCTTTAGCAAAATCTGGAAT 57
DB 73 GATGAATATCTCAACCATCTTGAGAGCACTTTGGCAATATCTAGCACT 24

Search completed: February 27, 2006, 07:58:48
Job time : 74.0816 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 38.5681 Seconds
(without alignments)
3538.457 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaagtgcttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA New.*

1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	27.2	42.5	554	US-09-925-065A-227036	Sequence 227036, A
C 2	27.2	42.5	556	US-09-925-065A-171598	Sequence 171598, A
C 3	26	40.6	201	US-10-995-561-31012	Sequence 31012, A
C 4	26	40.6	103660	US-10-995-561-13253	Sequence 13253, A
5	24.8	38.8	643	US-09-925-065A-897354	Sequence 897354, A
6	24.8	38.8	10302	US-11-127-832-23	Sequence 23, Appl
C 7	24.4	38.1	201	US-10-995-561-64267	Sequence 64267, A
C 8	24.4	38.1	201	US-10-995-561-64483	Sequence 64483, A
C 9	24.4	38.1	523	US-09-925-065A-312415	Sequence 312415, A
10	24.4	38.1	852	US-09-925-065A-59275	Sequence 59275, A
11	24.4	37.8	201	US-11-124-368A-11876	Sequence 11876, A
C 12	24.2	37.8	2416	US-09-925-065A-709999	Sequence 709999, A
C 13	24.2	37.8	2416	US-09-925-065A-710000	Sequence 710000, A
C 14	24.2	37.8	2416	US-09-925-065A-710001	Sequence 710001, A
C 15	24	37.5	201	US-10-995-561-64268	Sequence 64268, A
C 16	24	37.5	382	US-11-031-356-24	Sequence 24, Appl
C 17	24	37.5	519	US-09-925-065A-225239	Sequence 225239, A
18	24	37.5	584	US-09-925-065A-608992	Sequence 608992, A
19	24	37.5	2437	US-11-072-512-1159	Sequence 1159, Ap
C 20	24	37.5	56448	US-10-995-561-13369	Sequence 13369, A

21	23.8	37.2	201	12	US-11-124-368A-11820	Sequence 11820, A
22	23.8	37.2	569	6	US-09-925-065A-269497	Sequence 269497, A
23	23.8	37.2	569	6	US-09-925-065A-269498	Sequence 269498, A
C 24	23.8	37.2	1922	8	US-10-750-185-38496	Sequence 38496, A
C 25	23.8	37.2	1922	8	US-10-750-623-18496	Sequence 18496, A
26	23.6	36.9	741	6	US-09-925-065A-66253	Sequence 66253, A
C 27	23.6	36.9	1159	8	US-10-750-185-27344	Sequence 27344, A
C 28	23.6	36.9	1159	8	US-10-750-623-17344	Sequence 17344, A
29	23.6	36.9	1468	6	US-09-925-065A-698874	Sequence 698874, A
30	23.6	36.9	2064	12	US-11-117-169-7	Sequence 5, Appl
31	23.6	36.9	2186	12	US-11-117-169-7	Sequence 7, Appl
32	23.6	36.9	3068	12	US-11-117-169-14	Sequence 1, Appl
33	23.6	36.9	5032	8	US-10-775-169-75	Sequence 75, Appl
C 34	23.6	36.9	5032	8	US-10-947-249-37	Sequence 37, Appl
C 35	23.4	36.6	442	6	US-09-925-065A-586975	Sequence 586975, A
C 36	23.4	36.6	656	6	US-09-925-065A-671345	Sequence 671345, A
C 37	23.2	36.2	596	6	US-09-925-065A-308202	Sequence 308202, A
C 38	23.2	36.2	1318	6	US-09-925-065A-681453	Sequence 681453, A
C 39	23.2	36.2	1318	6	US-09-925-065A-681454	Sequence 681454, A
C 40	23.2	36.2	175100	12	US-11-121-086-21	Sequence 21, Appl
41	23	35.9	201	12	US-11-124-368A-11889	Sequence 11889, A
42	23	35.9	1167	6	US-09-925-065A-77828	Sequence 77828, A
43	23	35.9	1773	9	US-11-040-218-12	Sequence 12, Appl
C 44	23	35.9	1847	9	US-11-072-512-1670	Sequence 1670, Ap
45	23	35.9	2557	8	US-10-775-169-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-09-925-065A-227036/C
Sequence 227036, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925, 065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243, 096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252, 147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250, 092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261, 766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289, 846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 227036
LENGTH: 554
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-227036

Query Match 42.5%; Score 27.2; DB 6; Length 554;
Best Local Similarity 64.1%; Pred. No. 3.1;
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAAGGTTTCAACGCTGAGAGAGCTTTTACCAAAATCTGAACTGGA 60
DB GCTCCCAACAAGATTGTCAGGCTCATGAGATCTTGAACCAAAATTGCAATTGGA 237

QY 61 AGTA 64

DB 236 GGA 233

RESULT 2

```
US-09-925-065A-171598/C
; Sequence 171598, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108627.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 171598
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-171598
```

Query Match	42.5%	Score 27.2;	DB 6;	Length 556;
Best Local Similarity	64.1%;	Pred. No. 3.1;		
Matches	41;	Conservative	0;	Mismatches 23; Indels 0; Gaps 0;
QY	1	GATCCGCAAGAAGSTGTTCAAAAGCCCGGAGAGCTGTTTAGCAAAATCTGAACTGGA	60	
Db	296	GCTCCAAACAGATTGGTCAGAGCTCAATGAGAGAAATCTTTGAACCAAAATGCAAAATTGGA	237	
QY	61	AGTA	64	
Db	236	GGTA	233	

```

RESULT 3
US-10-995-561-31012/c
; Sequence 31012, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31012
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-31012

```

Query Match	Similarity	Score	DB	Length
Best Local	70.0%	Pred. No. 6.4		
Matches	35	Conservative	0	Mismatches 15; Indels 0; Gaps 0

Qy	8	GAAGAGGTGTTCAAA	CGCCTGGAGAGGCTTTT	AGCAAAATCTGAACT	57
db	73	GATGAAATGTTCAACCAAT	CTTTGAAGAGGACATCTT	GTGGSCATATCTGACCT	24

RESULT 4
US-10-995-561-13253/c
i Sequence 13253, Application US/10995561

```

? Publication No. US20050272054A1
? GENERAL INFORMATION:
? APPLICANT: CARGILL, Michele et al.
? TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
? TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
? TITLE OF INVENTION: DETECTION AND USES THEREOF
? FILE REFERENCE: CLO001559
? CURRENT APPLICATION NUMBER: US/10/995,561
? CURRENT FILING DATE: 2004-11-24
? NUMBER OF SEQ ID NOS: 85702
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 13253
? LENGTH: 103660
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(103660)
? OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-5)
US-10-995-561-13253

```

Query Match	40.6%	Score 26;	DB 8;	Length 103660;
Similarity	70.0%	Pred. No. 31;		
Best Local	35;	Conservative	0;	Mismatches 15; Indels 0; Gaps 0.

Oy	8	GAAGAGCGTGTTCAAACGCGCTGAGAGAGCTTTTACGAAATTCGAACT	57
db	47494	GATGAAAATGTTCACCATCTTTGAAGAGCACTTTGGCCAAATTTCTAGCACT	47445

```

RESULT: 5
US-09-925-065A-897354
; Sequence 897354, Application US/09925065A
; Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIORITY APPLICATION NUMBER: US 60/243,096
PRIORITY FILING DATE: 2000-10-24
PRIORITY APPLICATION NUMBER: US 60/252,147
PRIORITY FILING DATE: 2000-11-20
PRIORITY APPLICATION NUMBER: US 60/250,092
PRIORITY FILING DATE: 2000-11-30
PRIORITY APPLICATION NUMBER: US 60/261,766
PRIORITY FILING DATE: 2001-01-16
PRIORITY APPLICATION NUMBER: US 60/289,846
PRIORITY FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 897354
LENGTH: 643
TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-897354

```

	Query Match	38.8%	Score 24.8;	DB 6;	Length 643;
	Best Local Similarity	67.3%	Pred. No. 23;		
	Matches	35;	Mismatches	17;	Gaps 0
OY	9 AAGAAGTGTTCAAACGCTTGAGAAAGCTGTATTAGCAAAATCTGGAACTCGA	60			
DB	530 AACTAACATGTCATTAAGCTTGAGAAAGCTGTGCTGGGAAGCAGCAACTCGAA	581			

```

RESULT 6
US-11-127-832-23
; Sequence 23, Application US/11127832
; Publication No. US20060008884A1
; GENERAL INFORMATION:

```

APPLICANT: Hearing, Patrick
APPLICANT: Bahou, Madie
APPLICANT: Sandalon, Ziv
APPLICANT: Gnatelko, Dmitri
TITLE OF INVENTION: Adenoviral Vectors
FILE REFERENCE: STONY-04970
CURRENT APPLICATION NUMBER: US/11/127,832
CURRENT FILING DATE: 2005-05-12
PRIOR APPLICATION NUMBER: US/09/782,378
PRIOR FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/237,747
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Patencin version 3.0
SEQ ID NO 23
LENGTH: 10302
TYPE: DNA
ORGANISM: Homo sapiens
US-11-127-832-23

Query Match 38.8%; Score 24.8; DB 12; Length 10302;
Best Local Similarity 63.3%; Pred. No. 47;
Matches 38; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4 CCGCGAAGAGGTGTTCAACGCTTGAGAACTGTTGAACTGGAAGT 63
DB 2652 CCGCTACCAAGCTGTACAGAGGCTGTAGAGATCGTCAACACATCTAGAGATGAAGT 2711

RESULT 7
US-10-995-561-64267/c
Sequence 64267, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: CL001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 64267
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-64267

Query Match 38.1%; Score 24.4; DB 8; Length 201;
Best Local Similarity 68.0%; Pred. No. 24;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 9 AAGAAGTGTTCACAGCCCTGAGAGAGCTTTAGCAAAATCTGGAAGT 58
DB 135 AAGAAGATTGCTCAGGCAATGAGTAATGTCAACAAATCTGAATTTG 86

RESULT 8
US-10-995-561-64483/c
Sequence 64483, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: CL001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 64483

LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-64483

Query Match 38.1%; Score 24.4; DB 8; Length 201;
Best Local Similarity 68.0%; Pred. No. 24;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 9 AAGAAGTGTTCACAGCCCTGAGAGAGCTTTAGCAAAATCTGGAAGT 58
DB 64 AAGAAGATTGCTCAGGCAATGAGTAATGTCAACAAATCTGAATTTG 15

RESULT 9
US-09-925-065A-312415/c
Sequence 312415, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 312415
LENGTH: 523
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-312415

Query Match 38.1%; Score 24.4; DB 6; Length 523;
Best Local Similarity 68.0%; Pred. No. 31;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGGTGTTCAACGCTTGAGAGAGCTTTAGCAAAATC 50
DB 211 GAGATGTGGGGAAGGTCTCAAAAGCCAGAGACCATTTTGTGTAAGC 162

RESULT 10
US-09-925-065A-59275
Sequence 59275, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09

NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 59275
LENGTH: 852
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-59275

Query Match 38.1%; Score 24.4; DB 6; Length 852;
Best Local Similarity 63.8%; Pred. No. 35;
Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 4 CCGGAGAGAGAGTGTTCAAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 61
Db 663 CAGTGAATTAAGTGAATTAAGATTGATATACAGTTAGTGAATATAGAAATGGA 720

RESULT 11
US-11-124-368A-11876

Sequence 11876, Application US/11124368A
Publication No. US20050287559A1

GENERAL INFORMATION:
APPLICANT: Michele Cargili
APPLICANT: James J. Devlin
APPLICANT: May Luke

TITLE OF INVENTION: Genetic Polymorphisms Associated with
TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof

FILE REFERENCE: CU001524

CURRENT APPLICATION NUMBER: US/11/124.368A

PRIOR FILING DATE: 2005-05-09

PRIOR APPLICATION NUMBER: US 60/568,845

PRIOR FILING DATE: 2004-05-07

PRIOR APPLICATION NUMBER: US 60/625,936

PRIOR FILING DATE: 2004-11-09

NUMBER OF SEQ ID NOS: 21112

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11876

LENGTH: 201

TYPE: DNA

ORGANISM: Homo sapiens

US-11-124-368A-11876

Query Match 37.8%; Score 24.2; DB 12; Length 201;
Best Local Similarity 66.0%; Pred. No. 29;
Matches 35; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 11 GAAGGTGTTCAACCGCTGGAGAGCTGTTAGCAAAATCTGAACTGAACT 63
Db 2 GATGAGGTTCAGATCATGATGAGAAACAGAGGTAATTCTGATCTGAAAGT 54

RESULT 12

US-09-925-065A-709999/c

Sequence 709999, Application US/09925065A

Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single

TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

FILE REFERENCE: 108827.135

CURRENT APPLICATION NUMBER: US/09/925,065A

PRIOR FILING DATE: 2001-08-08

PRIOR APPLICATION NUMBER: US 60/243,096

PRIOR FILING DATE: 2000-10-24

PRIOR APPLICATION NUMBER: US 60/252,147

PRIOR FILING DATE: 2000-11-20

PRIOR APPLICATION NUMBER: US 60/250,092

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: US 60/261,766

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/289,846

PRIOR FILING DATE: 2001-05-09

NUMBER OF SEQ ID NOS: 957086

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 709999
LENGTH: 2416
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-709999

Query Match 37.8%; Score 24.2; DB 6; Length 2416;
Best Local Similarity 71.1%; Pred. No. 53;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 17 GTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 61
Db 878 GTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 834

RESULT 13

US-09-925-065A-710000/c

Sequence 710000, Application US/09925065A

Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single

TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

FILE REFERENCE: 108827.135

CURRENT APPLICATION NUMBER: US/09/925,065A

PRIOR FILING DATE: 2001-08-08

PRIOR APPLICATION NUMBER: US 60/243,096

PRIOR FILING DATE: 2000-10-24

PRIOR APPLICATION NUMBER: US 60/252,147

PRIOR FILING DATE: 2000-11-20

PRIOR APPLICATION NUMBER: US 60/250,092

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: US 60/261,766

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/289,846

PRIOR FILING DATE: 2001-05-09

NUMBER OF SEQ ID NOS: 957086

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 710000

LENGTH: 2416

TYPE: DNA

ORGANISM: Homo sapiens

US-09-925-065A-710000

Query Match 37.8%; Score 24.2; DB 6; Length 2416;
Best Local Similarity 71.1%; Pred. No. 53;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 17 GTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 61
Db 878 GTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 834

RESULT 14

US-09-925-065A-710001/c

Sequence 710001, Application US/09925065A

Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single

TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

FILE REFERENCE: 108827.135

CURRENT APPLICATION NUMBER: US/09/925,065A

PRIOR FILING DATE: 2001-08-08

PRIOR APPLICATION NUMBER: US 60/243,096

PRIOR FILING DATE: 2000-10-24

PRIOR APPLICATION NUMBER: US 60/252,147

PRIOR FILING DATE: 2000-11-20

PRIOR APPLICATION NUMBER: US 60/250,092

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: US 60/261,766

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/289,846
 PRIOR FILING DATE: 2001-05-09
 NUMBER OF SEQ ID NOS: 957086
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 710001
 LENGTH: 2416
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-925-065A-710001

Query Match 37.8%; Score 24.2; DB 6; Length 2416;
 Best Local Similarity 71.1%; Pred. No. 53;
 Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 17 GTTCAACGCTGTGAGAACCTGTTAGCAAAATCTGGAATGGA 61
 DB 878 GTTCAACGCTGTGAGAACCTGTTAGCAAAATCTGGAATGGA 834

RESULT 15
 US-10-995-561-64268/c
 Sequence 64268, Application US/10995561
 Publication No. US20050272054A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michele et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 TITLE OF INVENTION: DETECTION AND USES THEREOF
 FILE REFERENCE: CL001559
 CURRENT APPLICATION NUMBER: US/10/995,561
 CURRENT FILING DATE: 2004-11-24
 NUMBER OF SEQ ID NOS: 85702
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 64268
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-995-561-64268

Query Match 37.5%; Score 24; DB 8; Length 201;
 Best Local Similarity 66.0%; Pred. No. 34;
 Matches 33; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 9 AAGAAGTGTTCACACGCTGAGAACCTGTTAGCAAAATCTGGAATG 58
 DB 106 AAGAAGTGTTCACACGCTGAGAACCTGTTAGCAAAATCTGGAATG 57

Search completed: February 27, 2006, 11:38:29
 Job time : 38.5681 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 15.0601 Seconds
(without alignments) 7553.984 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaaggtgttc.....aaatctggaactgaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/1_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/H_COMB.seq: *
6: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq: *
7: /cgn2_6/prodata/1/ina/PP_COMB.seq: *
8: /cgn2_6/prodata/1/ina/RP_COMB.seq: *
9: /cgn2_6/prodata/1/ina/Backfillseq1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
C 1	26	40.6	2681 2	US-08-732-192A-1 Sequence 1, Appl
C 2	26	40.6	2681 3	US-09-172-339-7 Sequence 7, Appl
C 3	25.4	39.7	834 3	US-09-540-236-945 Sequence 945, App
4	25.4	39.7	269223 3	US-09-586-002-41 Sequence 41, App
5	25.2	39.4	601 3	US-09-949-016-25005 Sequence 25005, A
6	25.2	39.4	601 3	US-09-949-016-38480 Sequence 38480, A
7	25.2	39.4	601 3	US-09-949-016-170705 Sequence 170705, A
8	25.2	39.4	601 3	US-09-949-016-170748 Sequence 170748, A
9	25.2	39.4	601 3	US-09-949-016-170791 Sequence 170791, A
10	25.2	39.4	601 3	US-09-949-016-170832 Sequence 170832, A
11	25.2	39.4	601 3	US-09-949-016-176054 Sequence 176054, A
12	25.2	39.4	3416 3	US-09-949-016-4827 Sequence 4827, Ap
13	25.2	39.4	3416 3	US-09-949-016-4828 Sequence 4828, Ap
14	25.2	39.4	3490 3	US-09-949-016-4829 Sequence 4829, Ap
15	25.2	39.4	3490 3	US-09-949-016-4830 Sequence 4830, Ap
16	25.2	39.4	12241 3	US-09-949-016-16746 Sequence 16746, A
17	25.2	39.4	12243 3	US-09-949-016-12094 Sequence 12094, A
18	25.2	39.4	25230 3	US-09-949-016-13153 Sequence 13153, A
19	25.2	39.4	36075 3	US-09-949-016-16571 Sequence 16571, A
20	25.2	39.4	36075 3	US-09-949-016-16572 Sequence 16572, A
21	25.2	39.4	36625 3	US-09-949-016-12788 Sequence 12788, A
22	25.2	39.4	37133 3	US-09-949-016-16569 Sequence 16569, A
23	25.2	39.4	37133 3	US-09-949-016-16570 Sequence 16570, A
24	25	39.1	4174 3	US-09-949-002-138 Sequence 138, App

25	25	39.1	4180 3	US-09-814-915A-93 Sequence 93, Appl
25	25	39.1	4180 3	US-09-949-002-103 Sequence 103, App
26	25	39.1	15044 3	US-09-949-002-675 Sequence 675, App
27	25	39.1	15044 3	US-09-949-002-710 Sequence 710, App
28	25	39.1	15044 3	US-09-949-002-710 Sequence 710, App
29	24.8	38.8	601 3	US-09-949-016-25005 Sequence 25005, A
30	24.8	38.8	601 3	US-09-949-016-176055 Sequence 176055, A
31	24.8	38.8	10302 3	US-10-149-736-3 Sequence 3, Appl
32	24.8	38.8	10302 3	US-09-782-378A-23 Sequence 23, Appl
33	24.8	38.8	10320 3	US-09-091-501B-9 Sequence 9, Appl
34	24.6	38.4	3514 3	US-09-716-964B-117 Sequence 117, App
35	24.6	38.4	23456 3	US-09-949-016-12989 Sequence 12989, A
36	24.6	38.4	23456 3	US-09-949-016-12989 Sequence 12989, A
37	24.4	38.1	1853 3	US-08-687-590-61 Sequence 61, Appl
38	24.2	37.8	601 3	US-09-949-016-166961 Sequence 166961, A
39	24.2	37.8	601 3	US-09-949-016-166961 Sequence 166961, A
40	24.2	37.8	771 3	US-09-328-352-1830 Sequence 1830, Ap
41	24.2	37.8	30271 3	US-09-949-016-12796 Sequence 12796, A
42	24.2	37.8	30272 3	US-09-949-016-14006 Sequence 14006, A
43	24.2	37.8	33125 3	US-09-949-016-16436 Sequence 16436, A
44	24.2	37.8	33125 3	US-09-949-016-16437 Sequence 16437, A
45	24	37.5	382 3	US-09-284-782-24 Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-732-192A-1/c
; Sequence 1, Application US/08732192A
; Patent No. 5849526
; GENERAL INFORMATION:
; APPLICANT: Pichersky, Brian
; TITLE OF INVENTION: USE OF LINALLOL SYNTHASE IN GENETIC
; TITLE OF INVENTION: ENGINEERING OF SCENT PRODUCTION
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,192A
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,146
; FILING DATE: 25-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HIGHLANDER, STEVEN L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UMIC:015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2681 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..2637
; US-08-732-192A-1
Query Match : 40.6%; Score 26; DB 2; Length 2681;

RESULT 6
US-09-949-016-38480
; Sequence 38480, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38480
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-38480

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCGAAGAGCTGTTCAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 7
US-09-949-016-170705
; Sequence 170705, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170705
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170705

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCGAAGAGCTGTTCAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 8
US-09-949-016-170748
; Sequence 170748, Application US/09949016
; Patent No. 6812339

GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170748
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170748

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCGAAGAGCTGTTCAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 9
US-09-949-016-170791
; Sequence 170791, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170791
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170791

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCGAAGAGCTGTTCAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 10
US-09-949-016-170832
; Sequence 170832, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307

```
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 170832
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-170832

Query Match          39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1 GATCCGGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 46
DB 294 GAACCGTCTGAGCTGATCAATGCCAGAGAGCTGTATCTA 339

RESULT 11
US-09-949-016-176054
/ Sequence 176054, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 176054
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-176054

Query Match          39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 3 TCCGCGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 48
DB 225 TCCGGGTAGAGGATTTGTGCGCATGTATGATTCGATTAGAAAA 280

RESULT 12
US-09-949-016-4827
/ Sequence 4827, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
```

```
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 4827
/ LENGTH: 3416
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-4827

Query Match          39.4%; Score 25.2; DB 3; Length 3416;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1 GATCCGGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 46
DB 765 GAACCGTCTGAGCTGATCAATGCCAGAGAGCTGTATCTA 810

RESULT 13
US-09-949-016-4828
/ Sequence 4828, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 4828
/ LENGTH: 3416
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-4828

Query Match          39.4%; Score 25.2; DB 3; Length 3416;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1 GATCCGGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 46
DB 765 GAACCGTCTGAGCTGATCAATGCCAGAGAGCTGTATCTA 810

RESULT 14
US-09-949-016-4829
/ Sequence 4829, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
```

SEQ ID NO 4829
LENGTH: 3490
TYPE: DNA
ORGANISM: Human
US-09-949-016-4829

Query Match 39.4%; Score 25.2; DB 3; Length 3490;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGGTGTTCAACGCTGGAGAGCTGTTAGCAA 46
DB 765 GAACCTTCTGGAGCTGATCAATGCCAGAGAGCTGTATCTA 810

RESULT 15
US-09-949-016-4830
Sequence 4830; Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4830
LENGTH: 3490
TYPE: DNA
ORGANISM: Human
US-09-949-016-4830

Query Match 39.4%; Score 25.2; DB 3; Length 3490;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGGTGTTCAACGCTGGAGAGCTGTTAGCAA 46
DB 765 GAACCTTCTGGAGCTGATCAATGCCAGAGAGCTGTATCTA 810

Search completed: February 27, 2006, 06:46:23
Job time: 17.2601 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 351.203 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtgaagaactg.....aagctgcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_scs:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	31.6	51.8	235519	14	AC164258
C 2	31.6	51.8	243489	14	AC162622
C 3	31.4	51.5	441	15	AF317660
C 4	30.8	50.5	36748	12	AF043703
C 5	30.8	50.5	260699	14	AC006893
C 6	30.4	49.8	110000	13	AY653733_06
C 7	30.2	49.5	549	6	AX654322
C 8	30.2	49.5	110000	15	AK107375
C 9	30.2	49.5	110000	15	AP008213_061
C 10	30.2	49.5	126938	15	AP005779
C 11	30.2	49.5	128553	15	AP003931
C 12	29.8	48.9	414	2	PFALLERC
C 13	29.8	48.9	414	2	PFASAC
C 14	29.8	48.9	414	2	PFACBAC
C 15	29.8	48.9	604	1	DO016624
C 16	29.8	48.9	1338	5	AY825025
C 17	29.8	48.9	1349	5	AY825026
C 18	29.8	48.9	1428	2	PFAMMSABC

19	29.8	48.9	6409	2	PFAPMMMSA
20	29.6	48.5	1280	15	AK061230
21	29.6	48.5	1741	15	AK071110
22	29.6	48.5	12402	1	AE013431
23	29.6	48.5	110000	15	AP008216_196
24	29.6	48.5	135789	15	AC051634
25	29.6	48.5	300029	15	AE017113
26	29.4	48.2	607	15	AY096696
27	29.4	48.2	873	15	AB035444
28	29.4	48.2	874	15	AY086483
29	29.4	48.2	900	15	AY063997
30	29.4	48.2	107200	15	AC006551
31	29.4	48.2	170120	14	AC145728
32	29.2	47.9	38138	2	AF003148
33	29.2	47.5	330724	14	CEY67H2
34	28.8	47.2	2145	2	DIHSC90R
35	28.8	47.2	3720	13	S76368
36	28.8	47.2	43658	13	HSV3PGEN
37	28.8	47.2	112930	13	HSGEND
38	28.8	47.2	119958	14	AY914084
39	28.8	47.2	224093	14	AC156886
40	28.6	46.9	3039	5	CR942671
41	28.6	46.9	3042	5	BC060352
42	28.6	46.9	110000	1	CR848038_01
43	28.4	46.6	354	6	AR550464
44	28.4	46.6	476	15	AF317661
45	28.4	46.6	31853	2	CER44D12

ALIGNMENTS

RESULT 1
AC164258/c
LOCUS
DEFINITION
Bos taurus clone CH240-148E22, *** SEQUENCING IN PROGRESS ***, 28
unordered pieces.

ACCESSION
AC164258.2 GI:68300969
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Bos taurus (cow)
ORGANISM
Bos taurus

REFERENCE

AUTHORS

Muzny, D., Maric, M., Metzger, M., Lee, A., Adamson, S., Adams, C., Alder, J., Allen, C., Allen, R., Albrooks, S., Amin, A., Anguiano, D., Anylebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Genter, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dint, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C., Gabel, A., Garcia, R., Garcia, J., Garner, T., Garza, M., Gebregergels, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvery, P., Havlak, P., Hawes, A., Henderson, N., Hernandez, R., Hines, S., Hladik, S., Hume, J., Idlebird, D., Jackson, A., Hollins, B., Howells, S., Hulak, S., Hume, J., Johnson, R., Jolivet, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kovar, C., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C., L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensbuewa, L., Loulesed, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Mitosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Nwokilemeah, O., Okwou, G., Olarnpunaagoon, A., Pal, S., Parks, K.,
Pastermak, S., Paul, H., Perez, A., Perez, D., Pfannkuch, C.,
Plapper, F., Poldexler, A., Popovic, D., Primus, B., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Riley, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojce, A., Rose, R., Rose, R., Ruiz, S.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Sivatsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Staimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wodden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weise, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.

Unpublished
2 (bases 1 to 235519)

Direct Submission
Submitted (19-JUN-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 235519)
Cow Genome Sequencing Consortium.

Direct Submission
Submitted (01-JUN-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2005 this sequence version replaced gi:167972689.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: F10D
Center clone name: CH240-148E22
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 221304 bases at least Q40
Consensus quality: 224455 bases at least Q30
Consensus quality: 226977 bases at least Q20
Estimated insert size: 224509; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

1	1231: contig of 1231 bp in length
1232	2496: gap of 1265 bp in length
2497	15849: contig of 1335 bp in length
15850	18899: gap of 50 bp
15900	33519: contig of 17620 bp in length
33520	33569: gap of 50 bp
33570	47007: contig of 13438 bp in length
47008	47057: gap of 50 bp
47058	49714: contig of 2657 bp in length
49715	49765: gap of 50 bp
49766	56220: contig of 6456 bp in length
56221	56643: gap of 423 bp
56644	58196: contig of 1553 bp in length
58197	58297: gap of unknown length
58298	78107: contig of 19811 bp in length
78108	78157: gap of 50 bp
78158	101919: contig of 23762 bp in length
101920	101969: gap of 50 bp
101970	105635: contig of 3666 bp in length
105636	106018: gap of 383 bp
106019	107358: contig of 1340 bp in length
107359	107930: gap of 572 bp
107931	113033: contig of 5103 bp in length
113034	113097: gap of 64 bp
113098	116951: contig of 23854 bp in length
116952	137051: gap of unknown length
137052	139311: contig of 2260 bp in length
139312	139361: gap of 50 bp
139362	145341: contig of 5980 bp in length
145342	146014: gap of 673 bp
146015	186956: contig of 40942 bp in length
186957	187056: gap of unknown length
187057	189448: contig of 2392 bp in length
189449	189627: gap of 179 bp
189628	200786: contig of 1159 bp in length
200787	200886: gap of unknown length
200887	203015: contig of 2129 bp in length
203016	203490: gap of 475 bp
203491	209254: contig of 5764 bp in length
209255	209354: gap of unknown length
209355	210664: contig of 1210 bp in length
210665	210664: gap of unknown length
210665	211903: contig of 1139 bp in length
211904	211903: gap of unknown length
211904	213042: contig of 1139 bp in length
213043	213142: gap of unknown length
213143	214421: contig of 1279 bp in length
214422	214521: gap of unknown length
214522	216214: contig of 1693 bp in length
216215	216314: gap of unknown length
216315	217352: contig of 1038 bp in length
217353	217452: gap of unknown length
217453	219446: contig of 1994 bp in length
219447	219546: gap of unknown length
219547	235519: contig of 15973 bp in length.

Location/Qualifiers
1. .235519
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-148E22"
1232. .2496
/estimated_length=1265
15850. .15899
/estimated_length=50
33520. .33569
/estimated_length=50
47008. .47057
/estimated_length=50
49715. .49764
/estimated_length=50

*	25286	25335	gap of 50 bp
*	25336	33236	contig of 7901 bp in length
*	33337	33286	gap of 50 bp
*	33387	34642	contig of 1366 bp in length
*	34643	35251	gap of 609 bp
*	35252	37801	contig of 2550 bp in length
*	37802	38123	gap of 322 bp
*	38124	55860	contig of 17737 bp in length
*	55861	55960	gap of unknown length
*	55961	69472	contig of 13513 bp in length
*	69473	69523	gap of 50 bp
*	69524	74878	contig of 5355 bp in length
*	74879	75340	gap of 462 bp
*	75341	77411	contig of 2071 bp in length
*	77412	77511	gap of unknown length
*	77512	104543	contig of 27032 bp in length
*	104544	104931	gap of 386 bp
*	104932	10567	contig of 2636 bp in length
*	10568	10761	gap of 50 bp
*	107618	121698	contig of 14081 bp in length
*	121699	121748	gap of 50 bp
*	121749	133466	contig of 11718 bp in length
*	133467	133566	gap of unknown length
*	133567	138859	contig of 2293 bp in length
*	138860	135909	gap of 50 bp
*	135910	159070	contig of 23161 bp in length
*	159071	159120	gap of 50 bp
*	159121	176278	contig of 17158 bp in length
*	176279	176328	gap of 50 bp
*	176329	178842	contig of 3514 bp in length
*	178843	180440	gap of 596 bp
*	180441	209534	contig of 29094 bp in length
*	209535	209584	gap of 50 bp
*	209585	214425	contig of 4841 bp in length
*	214426	214475	gap of 50 bp
*	214476	222990	contig of 8515 bp in length
*	222991	223040	gap of 50 bp
*	223041	228305	contig of 5265 bp in length
*	228306	228435	gap of 130 bp
*	228436	231333	contig of 3688 bp in length
*	231334	232323	gap of unknown length
*	232324	232380	contig of 1047 bp in length
*	233381	233380	gap of unknown length
*	233381	233491	contig of 1011 bp in length
*	234392	234491	gap of unknown length
*	234492	235521	contig of 1030 bp in length
*	235522	235621	gap of unknown length
*	235622	236678	contig of 1057 bp in length
*	236679	236778	gap of unknown length
*	236780	238364	contig of 1566 bp in length
*	238365	238464	gap of unknown length
*	238465	239925	contig of 1461 bp in length
*	239926	240025	gap of unknown length
*	240026	241242	contig of 1217 bp in length
*	241243	241342	gap of unknown length
*	241343	243489	contig of 2147 bp in length

UNES	location/Qualifiers
source	1. .24369
	/organism="Bos taurus"
	/mol_type="genomic DNA"
	/db_xref="taxon:9913"
	/clone="CH240-125A15"
gap	11093. .11142
	/estimated_length=50
gap	12664. .12763
	/estimated_length=unknown
gap	25286. .25335
	/estimated_length=50
gap	33337. .33286
	/estimated_length=50
gap	34643. .35251
	/estimated_length=609
gap	37602. .38123

```
gap      /estimated_length=332
55861. .55960
gap      /estimated_length=unknown
69474. .69523
gap      /estimated_length=50
74879. .75340
gap      /estimated_length=462
77412. .77511
```

Query Match	51.8%	Score 31.6;	DB 14;	Length 243489;
Best Local Similarity	80.4%	Pred. No. 39;		
Matches 37; Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0;

Qy	11	GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA	56
Db	37145	GAAGAACTCAACAGAGAAAGTGATCAAGAGCTGGCCCAAGAA	37190

RESULT 3	AF317660	LOCUS	DEFINITION
	AF317660	441 bp	DNA linear
		605	acidic ribosomal protein type p1-B (p1B) gene,
			complete cds.

VERSION	AF317660.1	GI:11229037
KEYWORDS		
SOURCE	Candida albicans	

REFERENCE
AUTHORS
TITLE

Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida. 1 (bases 1 to 441)
Abramczyk D., Tchorzewski M. and Grankowski N.
Cloning, expression and purification of the acidic ribosomal protein from *Candida albicans*

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 441)
AUTHORS	Abdamezyk,D., Tchozewski,M. and Grankowski,N.
TITLE	Direct Submission
JOURNAL	Submitted (01-NOV-2000)
FEATURES	Curie-Skłodowska University, Department of Molecular Biology, Maria Curie-Skłodowska University, Akademicka 19, Lublin 20-033, Poland Location/Qualifiers

```

/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="COM 8215"
/db_xref="taxon:5476"
<8..>334
/gene="PIB"
<8..>334
/gene="PIB"
/product="60S acidic ribosomal protein type P1-B"
mRNA

```

CDS

```

/gene="PIB"
/note="type P1"
/codon_start=1
/translation_table=12
/product="60S acidic ribosomal protein type P1-B"
/protein_id="AAC33241.1"
/db_xref="GI:11229038"
/translation="MSTEEASVSYALLIADAEQRITSEKLLAIRKAGANVDQWADVY
PAKVEEKNLKELLSFAAALPASGAAGSAGSAGAGGAALAEFAAEBAEBAEESDDDM
GFGLEFD"

```

Query Match	51.5%	Score 31.4	DB 15	Length 441
Best Local Similarity	77.6%	Pred. NC. 74		
Matches 38	Conservative 0	Mismatches 11	Indels 0	Gaps 0

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAAAGTGTCTCAAGAAACTG 59
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 GAAGCTGTGCCGAAGAAGCCGCTGAGAAGAAAGCTGCTGAAGAATCTG 302


```
gene
/clone="W10C8"
/complement(56..6878)
/gene="pop-1"
/locus_tag="W10C8.2"
/complement(join(56..525,1465..2017,5068..5186,6707..6878))
/gene="pop-1"
/locus_tag="W10C8.2"
/strand_name="W10C8.2"
/notes="contains similarity to Pfam domain PF00505 (HMG
high mobility group) box); coded for by the following C.
elegans cDNAs: CKS85925, CKS87970, OSTF153D9_1,
OSTR153D9_1, U37532, OSTF30017H05"
/codon_start=1
/product="posterior pharynx defect protein 1"
/protein_id="AAK21504.1"
/db_xref="GI:13384579"
/db_xref="Wormbase:W10C8.2"
/translation="MADELDGKVFRRDEADDDPMISGETSEQOLADKKEAVNE
AELDAGRNPSIDVLKSAFPEKPEWSPSPVYGAALAAAKOHFENMAPIINRAGHPNNONG
ALRSPIMEPWSGSPPEPMPSPVYGAALAAAKOHFENMAPIINRAGHPNNONG
MPYPMHPSNAPQVNDRAAGCGGAKKDDHYKPLNAPFMKRNKALLETIGNEK
QSAELNKLGRHDLSEKQAKYFEWAKKDETHKEYPWASRKNVAWKKTKR
RDKSIPSENNQKCRAPGVNTEWCKFCRKKECYATDRSGSDITVDSQGRGT
SGAYSSSESPSPANAGIALTTOQQAAMHTLMQMRLGTTGASTHVPSPASSS
AGRSPIDANASDSVDEEEDQIDPTVMQOTHTMLMQESMCTI"
/complement(9460..9841)
/locus_tag="W10C8.3"
/complement(join(9460..9561,9737..9841))
/locus_tag="W10C8.3"
/standard_name="W10C8.3"
/notes="contains similarity to Hepatitis B virus SHBs;
TR:091164"
/codon_start=1
/product="Hypothetical protein W10C8.3"
/protein_id="AAK21501.2"
/db_xref="GI:26251556"
/db_xref="Wormbase:W10C8.3"
/translation="MESADSWLSPNPFQIRQARKMDEFSRCSLRATIALIVCAWM
LNNTCLSFNSFRNFFLLPQPL"
/locus_tag="W10C8.1"
/gene="ccb-2"
/locus_tag="W10C8.1"
/locus_tag="W10C8.1"
/standard_name="W10C8.1"
/notes="contains similarity to Pfam domains PF00774
(Dihydropyridine sensitive L-type calcium channel (Beta
subunit)), PF00018 (SH3 domain); coded for by the
following C. elegans cDNAs: OSTF081A3_1, OSTR081A3_1"
/codon_start=1
/product="Calcium channel, beta subunit protein 2"
/protein_id="AAK21500.1"
/db_xref="GI:13384575"
/db_xref="Wormbase:W10C8.1"

CDS
join(23308..23373,23547..23642,23698..23820,24632..25174,
25843..25974,26267..26380)
/gene="ccb-2"
/locus_tag="W10C8.1"
/standard_name="W10C8.1"
/notes="contains similarity to Pfam domains PF00774
(Dihydropyridine sensitive L-type calcium channel (Beta
subunit)), PF00018 (SH3 domain); coded for by the
following C. elegans cDNAs: OSTF081A3_1, OSTR081A3_1"
/codon_start=1
/product="Calcium channel, beta subunit protein 2"
/protein_id="AAK21500.1"
/db_xref="GI:13384575"
/db_xref="Wormbase:W10C8.1"

gene
/locus_tag="W10C8.1"
/gene="ccb-2"
/locus_tag="W10C8.1"
/standard_name="W10C8.1"
/notes="contains similarity to Pfam domains PF00774
(Dihydropyridine sensitive L-type calcium channel (Beta
subunit)), PF00018 (SH3 domain); coded for by the
following C. elegans cDNAs: OSTF081A3_1, OSTR081A3_1"
/codon_start=1
/product="Calcium channel, beta subunit protein 2"
/protein_id="AAK21500.1"
/db_xref="GI:13384575"
/db_xref="Wormbase:W10C8.1"

Query Match 50.5%; Score 30.8; DB 2; Length 36748;
Beet Local Similarity 70.7%; Pred. No. 78;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
```

```
RESULT 5
AC006893 26069 bp DNA linear HTG 26-FEB-1999
LOCUS Caenorhabditis elegans clone Y719, *** SEQUENCING IN PROGRESS ***,
DEFINITION 43 unordered pieces.
AC006893
AC006893.2 GI:4309906
```

```
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Mar 1, 1999 this sequence version replaced gi:4263455.
* NOTE: This is a "working draft" sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3303: contig of 3303 bp in length
3304 3312: gap of unknown length
3313 3319: contig of 2883 bp in length
6196 6204: gap of unknown length
6205 6205: contig of 2501 bp in length
8706 8714: gap of unknown length
8715 27303: contig of 18589 bp in length
27304 27312: gap of unknown length
27313 29898: contig of 2586 bp in length
29899 32748: gap of unknown length
32749 32757: gap of unknown length
32758 42435: contig of 9678 bp in length
42436 42444: gap of unknown length
42445 45447: contig of 3003 bp in length
45448 45456: gap of unknown length
45457 56578: contig of 11122 bp in length
56579 56587: gap of unknown length
56588 59294: contig of 2707 bp in length
59295 59303: gap of unknown length
59304 70656: contig of 11353 bp in length
70657 70665: gap of unknown length
70666 78234: contig of 7569 bp in length
78235 78243: gap of unknown length
78244 81287: contig of 3043 bp in length
81288 81295: gap of unknown length
81296 87907: contig of 6611 bp in length
87908 87916: gap of unknown length
87917 90315: contig of 2400 bp in length
90316 90324: gap of unknown length
90325 93415: contig of 3091 bp in length
93416 93424: gap of unknown length
93425 96332: contig of 2908 bp in length
96333 96341: gap of unknown length
96342 102339: contig of 5998 bp in length
102340 102348: gap of unknown length
102349 104720: contig of 2372 bp in length
104721 104729: gap of unknown length
104730 113259: contig of 8530 bp in length
113260 113268: gap of unknown length
113269 117369: contig of 4101 bp in length
117370 117378: gap of unknown length
117379 119526: contig of 2148 bp in length
119527 119535: gap of unknown length
119536 121798: contig of 2263 bp in length
121799 121807: gap of unknown length
121808 130291: contig of 8484 bp in length
130292 130300: gap of unknown length
130301 146526: contig of 16226 bp in length
```

146527 146535: gap of unknown length
146536 156661: contig of 10126 bp in length
156662 156670: gap of unknown length
156671 161833: contig of 5163 bp in length
161834 161842: gap of unknown length
161843 169581: contig of 7739 bp in length
169582 169590: gap of unknown length
169591 174622: contig of 5032 bp in length
174623 174631: gap of unknown length
174632 177362: contig of 2731 bp in length
177363 177371: gap of unknown length
177372 180479: contig of 3108 bp in length
180480 180488: gap of unknown length
180489 185112: contig of 4824 bp in length
185113 185321: gap of unknown length
185322 189706: contig of 4385 bp in length
189707 189715: gap of unknown length
189716 199829: contig of 10214 bp in length
199830 199938: gap of unknown length
199939 209300: contig of 9362 bp in length
209301 209309: gap of unknown length
209310 215261: contig of 5952 bp in length
215262 215269: gap of unknown length
215270 223804: contig of 8535 bp in length
223805 223812: gap of unknown length
223813 225776: contig of 1964 bp in length
225777 225784: gap of unknown length
225785 229087: contig of 3303 bp in length
229088 229095: gap of unknown length
229096 241882: contig of 12787 bp in length
241883 241890: gap of unknown length
241891 249775: contig of 7885 bp in length
249776 249783: gap of unknown length
249784 253545: contig of 3762 bp in length
253546 253553: gap of unknown length
253554 260699: contig of 7146 bp in length.

FEATURES
source 1. 260699 Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/db_xref="taxon:6239"
/clone="Y71F9"

gap /estimated_length=unknown
gap 6196. 6204 /estimated_length=unknown
gap 8706. 8714 /estimated_length=unknown
gap 27304. 27312 /estimated_length=unknown
gap 29899. 29907 /estimated_length=unknown
gap 32749. 32757 /estimated_length=unknown
gap 42436. 42444 /estimated_length=unknown
gap 45448. 45456 /estimated_length=unknown
gap 56579. 56587 /estimated_length=unknown
gap 59295. 59303 /estimated_length=unknown
gap 70657. 70665 /estimated_length=unknown
gap 78235. 78243 /estimated_length=unknown
gap 81287. 81295 /estimated_length=unknown
gap 87907. 87915 /estimated_length=unknown
gap 90316. 90324 /estimated_length=unknown
gap 93416. 93424 /estimated_length=unknown

gap /estimated_length=unknown
gap 96333. 96341 /estimated_length=unknown
gap 102340. 102348 /estimated_length=unknown
gap 104721. 104729 /estimated_length=unknown
gap 113260. 113268 /estimated_length=unknown
gap 117370. 117378 /estimated_length=unknown
gap 119527. 119535 /estimated_length=unknown
gap 121799. 121807 /estimated_length=unknown
gap 130292. 130300 /estimated_length=unknown
gap 146527. 146535 /estimated_length=unknown
gap 156662. 156670 /estimated_length=unknown
gap 161834. 161842 /estimated_length=unknown
gap 169582. 169590 /estimated_length=unknown
gap 174623. 174631 /estimated_length=unknown
gap 177363. 177371 /estimated_length=unknown
gap 180480. 180488 /estimated_length=unknown
gap 185313. 185321 /estimated_length=unknown
gap 189707. 189715 /estimated_length=unknown
gap 199930. 199938 /estimated_length=unknown
gap 209301. 209309 /estimated_length=unknown
gap 215262. 215269 /estimated_length=unknown
gap 223805. 223812 /estimated_length=unknown
gap 225777. 225784 /estimated_length=unknown
gap 229088. 229095 /estimated_length=unknown
gap 241883. 241890 /estimated_length=unknown

Query Match 50.5%; Score 30.8; DB 14; Length 260699;
Best Local Similarity 70.7%; Pred. No. 67;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTCAAGAACT 58
DB 200571 GAGCCAGTGAAGTGGCTGCTCAAGAACATCTTACCAAGACCTGCTGAGAACT 200628

RESULT 6
AY653733_06
Sequence split into 12 fragments LOCUS AY653733 Accession AY653733
WPCOMMENT
Fragment Name Begin End
AY653733_00 1 110000
AY653733_01 100001 210000
AY653733_02 200001 310000
AY653733_03 300001 410000
AY653733_04 400001 510000
AY653733_05 500001 610000
AY653733_06 600001 710000
AY653733_07 700001 810000
AY653733_08 800001 910000

AY653733.09 900001 1010000
 AY653733.10 1000001 1100000
 AY653733.11 1000001 1181404
 Continuation (7 of 12) of AY653733 from base 600001 (AY653733 Acanthamoeba polyphaga mit

Query Match 49.8% Score 30.4; DB 13; Length 110000;
 Best Local Similarity 71.4% Pred. No. 94;
 Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 5 CACGTGGAAGAACTCTCAAGAAACCGCTGCTCAAGAAAGCTCTCAAGAAACTCT 60
 Db 99868 CCAGTTCAAGAACTCTCAAGAAACCAAGTTCAAGAAAGTTCTCAAGAAACCACT 99923

RESULT 7
 LOCUS AK654322 AX654322 549 bp DNA linear PAT 22-MAR-2003
 DEFINITION Sequence 4192 from Patent WO03000898.
 ACCESSION AX654322
 VERSION AX654322.1 GI:29157136
 KEYWORDS
 SOURCE Oryza sativa
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
 AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
 Katagiri, F., Quan, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.
 TITLE Plant genes involved in defense against pathogens
 JOURNAL Patent: WO 03000898-A 4192 03-JAN-2003;
 Syngenta Participations AG (CH)

FEATURES
 source
 1. 549
 location/Qualifiers
 /organism="Oryza sativa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4530"

ORIGIN
 Query Match 49.5% Score 30.2; DB 6; Length 549;
 Best Local Similarity 74.5% Pred. No. 1.6e+02;
 Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 11 GAAGAAATGCTGCTCAAGAAACCGCTGCTCAAGAAAGCTGCTCAAGAAACTGTA 61
 Db 205 GATGAACCTGTTGAAGAGCGACCTCTCTGTGAAGCTATTGAAGAGCTGGA 255

RESULT 8
 LOCUS AK107375 AK107375 1188 bp mRNA linear PLN 24-JUL-2003
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:002-127-B08, full
 insert sequence.
 ACCESSION AK107375
 VERSION AK107375.1 GI:3292584
 KEYWORDS FLI CDNA; oligo capping.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
 AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
 Agrobiological Sciences Rice Full-Length cDNA Project Team;
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
 Kikuchi, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
 Ohneda, K., Shishiki, T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group; Otsu, Y., Murakami, K.,
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
 Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
 Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikura, J.,
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,

Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
 Hara, A., Hashidume, W., Hayatsu, N., Imocani, K., Ishii, Y., Itoh, M.,
 Kagawa, I., Kondo, S., Kono, H., Miyazaki, A., Oosato, N., Oka, Y.,
 Sato, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Yoshino, M. and Hayashizaki, Y.
 TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
 JOURNAL Science 301 (5611), 376-379 (2003)
 PUBMED 12869764

REFERENCE
 AUTHORS 2 (bases 1 to 1188)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
 Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T.,
 Hori, F., Hotta, I., Iida, Y., Iida, Y., Ikeda, R., Imamura, K.,
 Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
 Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
 Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
 Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
 Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
 Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
 Namiki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K.,
 Numasaki, R., Ohneda, E., Ono, M., Ohtsuki, K., Oka, M., Ooka, H.,
 Oosato, N., Oka, Y., Otsu, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K.,
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
 Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
 Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
 Toyota, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
 Yoshimura, A.

TITLE Direct Submission
 JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)
 COMMENT This clone is one of the 28K full-length cDNA clones from japonica
 rice.
 URL: http://cdna01.dna.affrc.go.jp/cDNA/
 NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
 Yamamoto, M.
 FAIS Genome Sequencing & Analysis Group: Otsu, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
 Mizuno, K., Narikawa, R., Nishikura, J., Oka, M., Ryu, R., Sugano, S.,
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imocani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Oosato, N.,
 Oka, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Waki, K.,
 Yasunishi, A. and Hayashizaki, Y.
 Location/Qualifiers
 1. 1188
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="002-127-B08"

ORIGIN

Query Match 49.5%; Score 30.2; DB 15; Length 1188;
Best Local Similarity 74.5%; Pred. No. 1.5e+02;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 11 GAGGAAGTCTGCAAGAAACCCCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 392 GATGAAGCTGTGAAGAGGACGCTCTCTGTAAGCTATTGAAGAGCTGGA 442

RESULT 9
AP008213_061/c
WPCOMMENT

Sequence split into 297 fragments LOCUS AP008213 Accession AP008213

Fragment Name	Begin	End
AP008213_000	1	110000
AP008213_001	100001	210000
AP008213_002	200001	310000
AP008213_003	300001	410000
AP008213_004	400001	510000
AP008213_005	500001	610000
AP008213_006	600001	710000
AP008213_007	700001	810000
AP008213_008	800001	910000
AP008213_009	900001	1010000
AP008213_010	1000001	1100000
AP008213_011	1100001	1210000
AP008213_012	1200001	1310000
AP008213_013	1300001	1410000
AP008213_014	1400001	1510000
AP008213_015	1500001	1610000
AP008213_016	1600001	1710000
AP008213_017	1700001	1810000
AP008213_018	1800001	1910000
AP008213_019	1900001	2010000
AP008213_020	2000001	2110000
AP008213_021	2100001	2210000
AP008213_022	2200001	2310000
AP008213_023	2300001	2410000
AP008213_024	2400001	2510000
AP008213_025	2500001	2610000
AP008213_026	2600001	2710000
AP008213_027	2700001	2810000
AP008213_028	2800001	2910000
AP008213_029	2900001	3010000
AP008213_030	3000001	3110000
AP008213_031	3100001	3210000
AP008213_032	3200001	3310000
AP008213_033	3300001	3410000
AP008213_034	3400001	3510000
AP008213_035	3500001	3610000
AP008213_036	3600001	3710000
AP008213_037	3700001	3810000
AP008213_038	3800001	3910000
AP008213_039	3900001	4010000
AP008213_040	4000001	4110000
AP008213_041	4100001	4210000
AP008213_042	4200001	4310000
AP008213_043	4300001	4410000
AP008213_044	4400001	4510000
AP008213_045	4500001	4610000
AP008213_046	4600001	4710000
AP008213_047	4700001	4810000
AP008213_048	4800001	4910000
AP008213_049	4900001	5010000
AP008213_050	5000001	5110000
AP008213_051	5100001	5210000
AP008213_052	5200001	5310000
AP008213_053	5300001	5410000
AP008213_054	5400001	5510000
AP008213_055	5500001	5610000
AP008213_056	5600001	5710000

AP008213_057	5700001	5810000
AP008213_058	5800001	5910000
AP008213_059	5900001	6010000
AP008213_060	6000001	6110000
AP008213_061	6100001	6210000
AP008213_062	6200001	6310000
AP008213_063	6300001	6410000
AP008213_064	6400001	6510000
AP008213_065	6500001	6610000
AP008213_066	6600001	6710000
AP008213_067	6700001	6810000
AP008213_068	6800001	6910000
AP008213_069	6900001	7010000
AP008213_070	7000001	7110000
AP008213_071	7100001	7210000
AP008213_072	7200001	7310000
AP008213_073	7300001	7410000
AP008213_074	7400001	7510000
AP008213_075	7500001	7610000
AP008213_076	7600001	7710000
AP008213_077	7700001	7810000
AP008213_078	7800001	7910000
AP008213_079	7900001	8010000
AP008213_080	8000001	8110000
AP008213_081	8100001	8210000
AP008213_082	8200001	8310000
AP008213_083	8300001	8410000
AP008213_084	8400001	8510000
AP008213_085	8500001	8610000
AP008213_086	8600001	8710000
AP008213_087	8700001	8810000
AP008213_088	8800001	8910000
AP008213_089	8900001	9010000
AP008213_090	9000001	9110000
AP008213_091	9100001	9210000
AP008213_092	9200001	9310000
AP008213_093	9300001	9410000
AP008213_094	9400001	9510000
AP008213_095	9500001	9610000
AP008213_096	9600001	9710000
AP008213_097	9700001	9810000
AP008213_098	9800001	9910000
AP008213_099	9900001	10010000
AP008213_100	10000001	10110000
AP008213_101	10100001	10210000
AP008213_102	10200001	10310000
AP008213_103	10300001	10410000
AP008213_104	10400001	10510000
AP008213_105	10500001	10610000
AP008213_106	10600001	10710000
AP008213_107	10700001	10810000
AP008213_108	10800001	10910000
AP008213_109	10900001	11010000
AP008213_110	11000001	11110000
AP008213_111	11100001	11210000
AP008213_112	11200001	11310000
AP008213_113	11300001	11410000
AP008213_114	11400001	11510000
AP008213_115	11500001	11610000
AP008213_116	11600001	11710000
AP008213_117	11700001	11810000
AP008213_118	11800001	11910000
AP008213_119	11900001	12010000
AP008213_120	12000001	12110000
AP008213_121	12100001	12210000
AP008213_122	12200001	12310000
AP008213_123	12300001	12410000
AP008213_124	12400001	12510000
AP008213_125	12500001	12610000
AP008213_126	12600001	12710000
AP008213_127	12700001	12810000
AP008213_128	12800001	12910000
AP008213_129	12900001	13010000

```

AP008213_130 13000001 13110000
AP008213_131 13100001 13210000
AP008213_132 13200001 13310000
AP008213_133 13300001 13410000
AP008213_134 13400001 13510000
AP008213_135 13500001 13610000
AP008213_136 13600001 13710000
AP008213_137 13700001 13810000
AP008213_138 13800001 13910000
AP008213_139 13900001 14010000
AP008213_140 14000001 14110000
AP008213_141 14100001 14210000
AP008213_142 14200001 14310000
AP008213_143 14300001 14410000
AP008213_144 14400001 14510000
AP008213_145 14500001 14610000
AP008213_146 14600001 14710000
AP008213_147 14700001 14810000
AP008213_148 14800001 14910000
AP008213_149 14900001 15010000
AP008213_150 15000001 15110000
AP008213_151 15100001 15210000
AP008213_152 15200001 15310000
AP008213_153 15300001 15410000
AP008213_154 15400001 15510000
AP008213_155 15500001 15610000
AP008213_156 15600001 15710000
AP008213_157 15700001 15810000
AP008213_158 15800001 15910000
AP008213_159 15900001 16010000
AP008213_160 16000001 16110000
AP008213_161 16100001 16210000
AP008213_162 16200001 16310000
AP008213_163 16300001 16410000
AP008213_164 16400001 16510000
AP008213_165 16500001 16610000
AP008213_166 16600001 16710000
AP008213_167 16700001 16810000
AP008213_168 16800001 16910000
AP008213_169 16900001 17010000
AP008213_170 17000001 17110000
AP008213_171 17100001 17210000
AP008213_172 17200001 17310000
AP008213_173 17300001 17410000
AP008213_174 17400001 17510000
AP008213_175 17500001 17610000
AP008213_176 17600001 17710000
AP008213_177 17700001 17810000
AP008213_178 17800001 17910000
AP008213_179 17900001 18010000
AP008213_180 18000001 18110000
AP008213_181 18100001 18210000
AP008213_182 18200001 18310000
AP008213_183 18300001 18410000
AP008213_184 18400001 18510000
AP008213_185 18500001 18610000
AP008213_186 18600001 18710000
AP008213_187 18700001 18810000
AP008213_188 18800001 18910000
AP008213_189 18900001 19010000
AP008213_190 19000001 19110000
AP008213_191 19100001 19210000
AP008213_192 19200001 19310000
AP008213_193 19300001 19410000
AP008213_194 19400001 19510000
AP008213_195 19500001 19610000
AP008213_196 19600001 19710000

```

```

Query Match 49.5%; Score 30.2; DB 15; Length 110000;
Best Local Similarity 74.5%; Pred. No. 1.1e+02;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

11 GAAGAACTGCTCAAGAAACGCTGCTCAAGAAAGCTGCTCAAGAAACTGTA 61

```

```

Db 71010 GATGAACCTGTGAAGAGCAGCTGCTGTGAAGACTATGAAAGCTGGA 70960
RESULT 10
AP005779/c 126938 bp DNA linear PLN 22-JUL-2004
LOCUS Oryza sativa [japonica cultivar-group] genomic DNA, chromosome 7,
DEFINITION BAC clone:OSJNB0042J07.
ACCESSION AP005779
VERSION AP005779.3 GI:50508900
SOURCE Oryza sativa [japonica cultivar-group]
ORGANISM Oryza sativa [japonica cultivar-group]
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:OSJNB0042J07
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 126938)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasak@nias.affrc.go.jp, URL:http://xgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Jul 22, 2004 this sequence version replaced gi:33980494.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), Genemark.hmm
(http://opal.biology.gatech.edu/Genemark/), glimmerM
(http://www.tigr.org/tcb/glimmerM/glmr_form.html), RiceHMM
(http://xgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
(http://bioinformatics.iaestate.edu/cgi-bin/sp.cgi), sim4
(http://globin.cse.psu.edu/html/doc/sim4.html), gap2
(http://www.tigr.org/software/glimmerM/), BLASTN and BLASTX. The
genomic sequence was searched against NCBI NonRedundant Protein
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
sequence database at RGP or DBJ. Protein homologies of the coding
regions were searched against NCBI Nonredundant Protein database
with BLASTP. ESTs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no. and RGP clone ID.
Full-length cDNAs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with full-length cDNA or
EST homology (covering almost the entire length of partial
sequence) is classified as an 'unknown' protein. A gene predicted
by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from M13rev to -21M13 of the BAC
clone. This sequence of OSJNB0042J07 clone has an overlap with
OSJNB00008C11 (DBJ: AP005098) clone at 5' end and with OJ1664 D08
(DBJ: AP003931) at 3' end. Detailed information on overlap and
assembly quality together with annotation of this entry is
available at http://xgp.dna.affrc.go.jp/GenomesSeq.html.
Location/Qualifiers
1. 126938
/organism="Oryza sativa [japonica cultivar-group]"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="OSJNB0042J07"

```


DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7, BAC clone:OJ1664_D08.
ACCESSION	AP003931
VERSION	AP003931.3
KEYWORDS	GI:50508415
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE	1 Sasaki, T., Matsumoto, T. and Yamamoto, K. Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC clone:OJ1664_D08
AUTHORS	Published Only in Database (2001)
JOURNAL	2 (bases 1 to 128553) Sasaki, T., Matsumoto, T. and Yamamoto, K. Direct Submission Submitted (18-JUL-2001) Takuji Sasaki, National Institute of Agricultural Sciences, Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki, 305-8602, Japan (E-mail: tsasaki@nias.affrc.go.jp, URL: http://tsgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468) On Jul 22, 2004 this sequence version replaced gi:22531714. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), Genemark.hmm (http://opal.biology.gatech.edu/genemark/), GlimmerM (http://opal.biology.gatech.edu/genemark/), GlimmerM (http://www.tigr.org/tcb/glimmer/glmr_form.html), RiceHMM (http://tsgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://gljlab.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DBJ. Protein homologues of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hybrid', protein prediction programs is classified as a 'hybrid', protein prediction programs is classified as a 'hybrid', protein prediction programs is also classified as a predicted by a single gene prediction program is also classified as a probable 'hybrid' protein and is included as a miscellaneous feature of the sequence.
FEATURES	The orientation of the sequence is from -21M3 to M13rev of the BAC clone. This sequence of OJ1664_D08 clone has an overlap with OSUNB0042J07(DBJ: AP005779) at 5' end and with OJ116_C08 (DBJ: AP004002) at 3' end. The sequence was generated by combining Monbanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://tsgp.dna.affrc.go.jp/genomeseg.html.
SOURCE	Location/Qualifiers 1..128553 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="7" /clone="OJ1664_D08" /complement(join(1360..1450,1834..1870,1981..2011,3176..3208)) /gene="OJ1664_D08.101" complement(join(1360..1450,1834..1870,1981..2011,3176..3208)) /gene="OJ1664_D08.101"
misc_feature	
gene	
misc_feature	
gene	
mRNA	
gene	
mRNA	
gene	
misc_feature	
CDS	
mRNA	
gene	
CDS	

RDYCKRRRPFDDMAOKYIHOTAGILPEDRNDGARPDPMAERKAEMKLEISMELE
EMVYKOCVREKESLSLRTNLPFLFRVSPFSLASAMFGWGSIAHTNQDLE"
gene 30227..36900
/gene="OJ1664.D08.105"
30227..30585
/gene="OJ1664.D08.105"
/note="5, terminal repeat"
/rpt_type="terminal"
misc_feature 31648..36161
/gene="OJ1664.D08.105"
/note="probably inactive due to including stop codon(s) in
CDS
pseudogene, copia polypeptide"
repeat_unit 36539..36900
/gene="OJ1664.D08.105"
/note="3, terminal repeat"
/rpt_type="terminal"
gene 40221..40289
/gene="OJ1664.D08.106"
complement(join(39517..39634,39757..39812,39934..39987,
40221..40289))
/gene="OJ1664.D08.106"
/note="hypothetical ORF
predicted by GeneMark.hmm
this category is not included in IRGSP standard"
40722..43447
/gene="OJ1664.D08.107"
join(40722..41153,41226..41350,41434..41536,41644..41782,
42196..42473,43003..43447)
/gene="OJ1664.D08.107"
/note="supported by full-length cDNA(s): AK071575"
join(40723..41153,41226..41350,41434..41536,41644..41782,
42196..42473,43003..43438)
/gene="OJ1664.D08.107"
/note="supported by full-length cDNA(s): AK104057"
join(40785..41153,41226..41350,41434..41536,41644..41782,
42196..42473,43003..43167)
/gene="OJ1664.D08.107"
/note="contains EST(s): D48305(S14442), AU096613(S14442)
contains full-length cDNA(s): AK071575, AK104057"
/codon_start=1
/product="putative mRNA binding protein precursor"
/protein_id="BAC83225.1"
/db_xref="GI:34393296"
/translation="MAFSSSTAATAAPAFVSGHRRRGACHLSSFLIAFPVAAR
CEMAAAPPRAVAVTTRAGAAAGAGCKSVLTNTGCGHAAVIGFYAKDLAAGHVT
VITVDEGSDKKKPPFRSELTSGATTTWGDPAVCAAGGASFPVITDNRKQ
IDAVKPVVMAKAAGVAGFLFVSSAGITVPDSBPPEVDEDAVESGHTVERKIAEQ
FGSMASFPOYMGISGNKDCBEMFEDRIVRGSPVPIPGSGMQVTNISHVRLASVVA
LAVESPGAAGRIENCVDRAVTFNGLVWCAAAAOPEILHYDPAAGVADKAKFP
FNNMHPYAPRAKEVLTGWRSSTNLPEDLKEKRAEVAASSGRGCKEMSFLLDKITIAA

Query Match 49.5%; Score 30.2; DB 15; Length 128553;
Best Local Similarity 74.5%; Pred. No. 1.1e+02;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
DB 45565 GATGAACCTGTTGAAGAGCAGCTGCTCGTGAAGCTATTGAAGAACTGGA 45515

RESULT 12
LOCUS PF111ERC 414 bp DNA linear INV 26-Apr-1993
DEFINITION P.chabaudi merozoite surface antigen 1 (MSA-1) gene, partial cds.
ACCESSION M61208
VERSION M61208.1 GI:160022
KEYWORDS merozoite surface antigen 1.
SOURCE Plasmodium chabaudi
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 414)
AUTHORS Deleersnijder, W., Hendrix, D. and Hamers, R.
TITLE Analysis of MSA-1 diversity in Plasmodium chabaudi strains
JOURNAL Mol. Biochem. Parasitol. 46 (2), 315-317 (1991)
COMMENT 1922203
FEATURES
source
1..414
/organism="Plasmodium chabaudi"
/mol_type="genomic DNA"
/strain="11ER"
/db_xref="taxon:5825"
1..414
/gene="MSA-1"
1..414
/gene="MSA-1"
/codon_start=3
/product="merozoite surface antigen 1"
/protein_id="AAA29494.1"
/db_xref="GI:552183"
/translation="APTPAVPAPETAPETAPETAPETAPETAPETAPETAPET
TPPASASTETPPKAPATPTPTSETVTOEGTTPAAPKAGASSAPAPAPAKPAPAO
TVTGGSTNVEGSTVRAESDEMFVDDEVDFYK"

Query Match 48.9%; Score 29.8; DB 2; Length 414;
Best Local Similarity 75.5%; Pred. No. 2.2e+02;
Matches 37; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

ORIGIN

QY 13 AGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
DB 53 AGAACTGACCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 101

RESULT 13
LOCUS PFASAC 414 bp DNA linear INV 26-Apr-1993
DEFINITION P.chabaudi merozoite surface antigen 1 (MSA-1) gene, partial cds.
ACCESSION M61206
VERSION M61206.1 GI:160099
KEYWORDS merozoite surface antigen 1.
SOURCE Plasmodium chabaudi
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 414)
AUTHORS Deleersnijder, W., Hendrix, D. and Hamers, R.
TITLE Analysis of MSA-1 diversity in Plasmodium chabaudi strains
JOURNAL Mol. Biochem. Parasitol. 46 (2), 315-317 (1991)
COMMENT 1922203
FEATURES
source
1..414
/organism="Plasmodium chabaudi"
/mol_type="genomic DNA"
/strain="AS"
/sub_strain="chabaudi"
/db_xref="taxon:5825"
1..414
/gene="MSA-1"
1..414
/gene="MSA-1"
/codon_start=3
/product="merozoite surface antigen 1"
/protein_id="AAA29494.1"
/db_xref="GI:552183"
/translation="APTPAVPAPETAPETAPETAPETAPETAPETAPETAPET
TPPASASTETPPKAPATPTPTSETVTOEGTTPAAPKAGASSAPAPAPAKPAPAO
TVTGGSTNVEGSTVRAESDEMFVDDEVDFYK"

